2014

Modeling, inference and clustering for equivalence classes of 3-D orientations

Chuanlong Du
Iowa State University

Modeling, inference and clustering for equivalence classes of 3-D orientations

by

Chuanlong Du

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Stephen Vardeman, Co-major Professor

Daniel Nordman, Co-major Professor

Dan Nettleton

Huaiqing Wu

Guang Song

Iowa State University

Ames, Iowa

2014

DEDICATION

I would like to dedicate this dissertation to my mother Jinyan Xu. She always encourages me to follow my heart and to seek my dreams. Her words have always inspired me and encouraged me to face difficulties and challenges I came across in my life. I would also like to dedicate this dissertation to my wife Lisha Li without whose support and help I would not have been able to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this dissertation. First and foremost, Dr. Steve Vardeman and Dr. Dan Nordman for their guidance, patience and support throughout this research and the writing of this dissertation. Their insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. Second, I'd like to thank Dr. Dan Nettleton for supporting me as a research assistant during my last 3 year at Iowa State University. I would also like to thank my committee members for their efforts and contributions to this work.

ABSTRACT

Investigating cubic crystalline structures of specimens is an important way to study properties of materials in text analysis. Crystals in metal specimens have internally homogeneous orientations relative to a pre-chosen reference coordinate system. Clusters of crystals in the metal with locally similar orientations constitute so-called "grains." The nature of these grains (shape, size, etc.) affects physical properties (e.g., hardness, conductivity, etc.) of the material. Electron backscatter diffraction (EBSD) machines are often use to measure orientations of crystals in metal specimens. However, orientations reported by EBSD machines are in truth *equivalence classes* of crystallographically symmetric orientations.

Motivated by the materials science applications, we formulate parametric probability models for "unlabeled orientation data." This amounts to developing models on equivalence classes of 3-D rotations. A Bayesian method is developed for inferencing parameters in the models, which is generally superior to large-sample methods based on likelihood estimation. We also proposed an algorithms for clustering equivalence classes of 3-D orientations. As we continue to work on this area, we found and studied an interesting class of Markov chains with state spaces partitions of a finite set. These Markov chains have some properties that make them attractive in their own right, and they are potentially helpful in Bayesian model-based clustering.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

Texture analysis in materials science often involves examining cubic crystalline structures on a specimen's surface. Electron backscatter diffraction (EBSD) provides a method for collecting data on microtexture of metals, where data points indicate the orientations of cubic crystals on the surface of a metal specimen; see Randle (2003). Such measured crystallographic (cubic) orientations are typically reported as 3×3 rotation matrices in $SO(3)$ (i.e., the set of 3×3 orthogonal matrices with determinant 1). As illustrated in Figure 1.1, each column of a rotation matrix $\boldsymbol{O}$ represents where the corresponding columns of the 3×3 identity matrix $\boldsymbol{I}_3$ move upon rotation and thereby represents the orientation of edges of a cube after rotating its reference frame in $\mathbb{R}^3$ away from some "world" reference frame (given by the standard coordinate vectors in $\boldsymbol{I}_3$). However, a challenge in interpreting EBSD data is that there is an inherent ambiguity about which edges of the cubic crystal are being represented by the reported orientation matrices. That is, one actually observes only an alignment of a crystal cube *up to a labeling of the cube's sides*, and 24 possible labelings exist, all of which produce morphologically indistinguishable alignments relative to the coordinate system. To understand these potential labelings, note that in Figure 1.1 there are 8 ways to assign a corner of the cube to be the origin along with 3 ways to select an edge as the z axis (subsequently preserving the right-hand rule). Figure 1.2 shows a sample of 6 such labelings (from the 24 coordinate systems) based on selecting the 2 bottom left corners as the origins. So, a measurement from EBSD does not truly represent a single orientation in the usual sense, but rather an equivalence class consisting of 24 rotations (of which the reported orientation is an element), all of which are equally valid descriptions of a cube's alignment relative to the coordinate system. We henceforth refer to

such observations as *unlabeled* orientation data, to distinguish them from true orientation data (i.e., data as rotations).

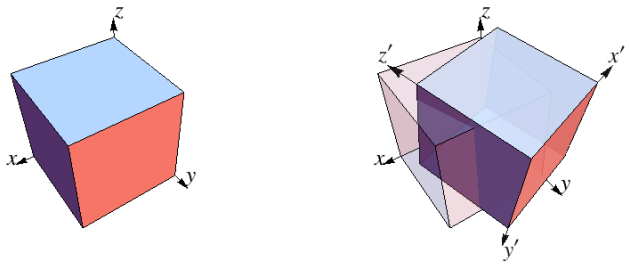

Figure 1.1 On the left, three edges of a (non-rotated) cube are aligned with a standard coordinate reference frame in $\mathbb{R}^3$; each individual edge has a direction corresponding to a column of 3×3 identity matrix $\boldsymbol{I}_3$. The right figure illustrates a cube after a rotation; the resulting orientation can be represented by a 3×3 rotation matrix $\boldsymbol{O} = [\boldsymbol{x}' \, \boldsymbol{y}' \, \boldsymbol{z}']$, where the columns of $\boldsymbol{O}$ indicate how columns of $\boldsymbol{I}_3$ (i.e., edges of the cube) move upon rotation. (Further, in the Euler axis-angle representation of a rotation from Appendix A, the rotation can be described by "turning" the cube through an angle r about a fixed axis in the direction of $\boldsymbol{u} \in \mathbb{R}^3$, $\|\boldsymbol{u}\| = 1$.)

Statistical tools for parametric inference with such unlabeled orientation data are potentially useful. However, there has been little development in this area outside of some early modeling work by Mackenzie and Thomson (1957), Handscomb (1958), Mackenzie (1958), Morawiec (1995), Morawiec (1996), Basson (1997), Morawiec (1998) and Basson (1998) who considered the distribution of so-called "random disorientation" for the alignment of an unlabeled cube in 3-dimensional space. Later, Morawiec (2004) covers wide and comprehensive topics about rotation and orientations. More recently, in the materials science literature, Bachmann et al. (2010) developed large-sample results for non-parametric inference with unlabeled orientation data, using the Bingham distribution on 4-dimensional quaternions (Bingham, 1974; Kunze and Schaeben, 2004, 2005), and Hielscher et al. (2010) proposed a least-squares-like method and non-parametric density estimator for the distribution of orientations of (unlabeled) cubic crystals within a single crystal grain. Neither of these works addresses parametric statistical inference for such data. Our first goal is to introduce a broad and interpretable class of para-

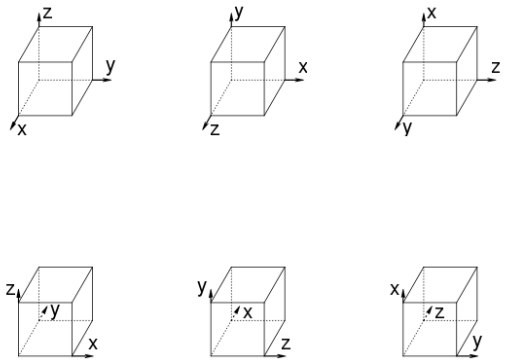

Figure 1.2 An illustration of 6 labelings (coordinate systems) on a cuboid based on selecting the 2 bottom left points as the origin.

metric probability models for unlabeled orientations and provide Bayesian tools for inference in these models. Grains maps (see Chapter 3.1) are commonly used to study materials in textual analysis. As we have argued here and will further demonstrate in Chapter 2.1 that EBSD measurements are in truth equivalence classes of orientations, we propose an hierarchical clustering method for producing grains maps based on equivalence classes of orientations with spatial information.

1.2 Dissertation Organization

This dissertation is a combined work of 3 papers that I worked on. All of these 3 papers are motivated or related to materials science applications. In the first paper (Chapter 2), We formulate parametric probability models for equivalence classes of 3-D orientations and develop one-sample Bayes inference for the parameters in these models. In the second paper (Chapter 3), we proposes a clustering-smoothing algorithm for producing grain maps for spatially informed equivalence classes of 3-D orientations. In the last paper (Chapter 4), we studies a class of Markov chains with state spaces partitions of a finite set and illustrate the potential use of these Markov chains in Bayesian model-based clustering.

CHAPTER 2. BAYESIAN INFERENCE FOR A NEW CLASS OF DISTRIBUTIONS ON EQUIVALENCE CLASSES OF 3-D ORIENTATIONS WITH APPLICATIONS TO MATERIALS SCIENCE

A paper tentatively accepted by Technometrics

Chuanlong Du [1], Daniel Nordman[2] and Stephen Vardeman [3]

Abstract

Experiments in materials science investigating cubic crystalline structures often collect data which are in truth *equivalence classes* of crystallographically symmetric orientations. These intend to represent how lattice structures of particles are orientated relative to a reference coordinate system. Motivated by a materials science application, we formulate parametric probability models for "unlabeled orientation data." This amounts to developing models on equivalence classes of 3-D rotations. We use a flexible existing model class for random rotations (called uniform-axis-random-spin models) to induce probability distributions on the equivalence classes of rotations. We develop one-sample Bayes inference for the parameters in these models, and compare this methodology to some likelihood-based approaches. We also contrast the new parametric analysis of unlabeled orientation data with other analyses that proceed as if the data have been pre-processed into honest orientation data.

[1]Primary researcher and author, graduate student, Department of Statistics, Iowa State University.
[2]Associate Professor, Department of Statistics, Iowa State University.
[3]University Professor, Department of Statistics, Department of Industrial and Manufacturing Systems Engineering, Iowa State University.

2.1 Introduction

To motivate our work and give a concrete illustration, we consider a data set collected in an investigation intended to quantify the precision of EBSD measurements. (This particular experiment involved repeatedly switching an EBSD device on/off to assess machine precision in output. Standard subject matter methodology is more usually based on one machine run, so that we do not suggest that this particular design has wide applicability. However, the data from this experiment are particularly helpful for illustrating and clarifying the labeling ambiguity in EBSD measurements.) The same region of the surface of a nickel specimen was scanned (at 3,449 locations) over 14 separate runs (i.e., the machine was turned off and back on 13 times and attempt was made to manually re-register the machine with a corner of the specimen on each repetition). For the 14 rotation matrices produced corresponding to a single location, the rotation angles between each of the $\binom{14}{2} = 91$ resulting pairs of rotation matrices was calculated. (A rotation angle $r \in [0, \pi]$ between two rotation matrices $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ is the angle needed to align one rotation matrix $\boldsymbol{O}_1$ back to another $\boldsymbol{O}_2$, given by $1 + 2\cos(r) = \text{trace}(\boldsymbol{M}(\boldsymbol{u}, r))$ for $\boldsymbol{M}(\boldsymbol{u}, r) = \boldsymbol{O}_1^T \boldsymbol{O}_2$ expressed in (A.1) of Appendix A.) Histograms of these rotation angles at one arbitrarily chosen location are presented in Figure 2.1 under the heading "before preprocessing." We can see that the rotation angles concentrate around 0, $\frac{\pi}{2}$ and π, indicating that the representations of essentially the same cubic orientation produced by the EBSD machine in different scans at the same location were not consistent, and differ due to ambiguity associated with crystallographic symmetries. For (pre-processed version of) these data, Bingham et al. (2009a) developed quasi-likelihood inference for rotation matrices (i.e., actual orientations) and later also developed Bayes inference methods (Bingham et al., 2009b, 2010). To plausibly treat EBSD measurements as rotation matrices, Bingham et al. (2009a) needed to attempt to remove the inconsistencies in representations evident in the repeat scans. The preprocessing was done by aligning each rotation matrix to the rotation matrix among its 24-fold equivalence class (see Chapter 2.2.1) having the minimal rotation angle between it and a pre-chosen/fixed rotation matrix from the 14 scans. The rotation angles between rotation matrices after preprocessing are shown in Figure 2.1 under the heading

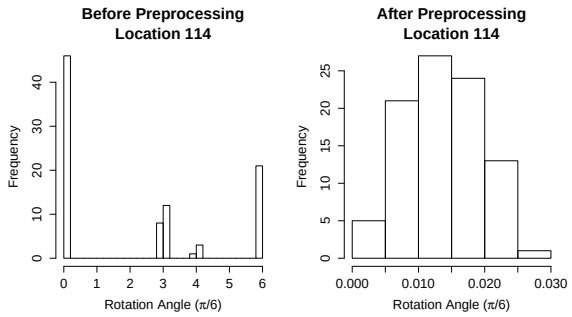

Figure 2.1 Histograms of rotation angles at locations 114, where the angles are expressed as multiples of $\pi/6$ showing concentrations at $(0, 3, 4, 6) * \pi/6 = (0, pi/2, pi, 2pi/3)$ (as expected under cube labeling ambiguities).

"after preprocessing." We can see that the rotation angles after processing are very close to 0, indicating that the EBSD data after preprocessing are highly concentrated, which suggests that this naive preprocessing works well in this situation. However, the preprocessing might work far less well if one had unlabeled orientation data that were not so highly concentrated, and then faced greater uncertainty concerning selection of an appropriate orientation representation for each data point from its 24-fold equivalence class of rotations (see Chapter 2.2.1). It thus seems preferable to conduct inference in a rigorous way that acknowledges the unlabeled nature of such data. (The topic of analyzing EBSD data as unlabeled orientations vs. "labeled" orientations after pre-processing will be examined later; see Chapter 2.4.)

We here use the uniform-axis-random-spin (UARS) models for rotation matrices from Bingham et al. (2009a) as a basis for developing general parametric probability models for equivalence classes of rotations. UARS models are useful for describing orientation data which arise from distributionally symmetric random perturbations of a fixed, mean rotation. (This is analogous to a location model for real-valued data that involves symmetric errors added to an underlying mean parameter.) Such models for random rotations are based on a simple geometric construction in terms of Euler's axis-angle representation of rotations (León et al., 2006, sec. 5.2; Bingham et al., 2009a; Hielscher et al., 2010) and are equivalent to rotationally symmetric distributions on quaternions mapped to $SO(3)$ (Watson, 1983; Prentice, 1986). The UARS model class is broad and includes many popular symmetric parametric models

such as the symmetric (or isotropic) matrix Fisher (Downs, 1972), Cayley (León et al., 2006) and isotropic Gaussian distributions (Savyolova and Nikolayev, 1995) on $SO(3)$. Because of this, it becomes natural to consider this model class as a basis for developing distributions for unlabeled orientation data. See the appendix for additional details of UARS modeling.

After introducing these new models for unlabeled rotations and providing some brief consideration of likelihood inference in Chapter 2.2, Chapter 2.3 then develops a one-sample Bayes inference method for relatively non-informative priors for the parameters in these new models on equivalence classes. We use Bayes inference for practical (rather than philosophical) reasons, because computations involved are very straightforward compared to a maximum likelihood approach (which requires reparameterizations and optimization over complicated manifolds as parameter spaces, and can fail to be defined in non-regular models). The Bayes approach also allows direct construction of geometrically interpretable credible regions. Through simulation, we compare our Bayes inference method to likelihood-based methods and show that (for at least one important model choice) our Bayes method outperforms them in terms of producing coverage rates matching nominal, especially when sample size is small. In Chapter 2.4, we return to the analysis of EBSD data and show by comparisons based on real and simulated data that (for two choices of model class) our Bayes inference method for unlabeled orientations is comparable to the Bayes inference method based on orientations derived from preprocessing when data are highly concentrated, and that our method is clearly superior when underlying orientation data are not highly concentrated. Chapter 2.5 then briefly describes extension of our work for modeling cubic orientations that are potentially "partially labeled," which concerns cases where some labeling ambiguity may exist but some sides/edges of an oriented object (e.g., a hyper-rectangle) may be distinguishable from each other. Chapter 2.6 then provides some concluding remarks.

We end this extended introduction by mentioning that considerably more literature exists, both in statistics and materials science, concerning modeling and inference for orientation data (rotation matrices). The most commonly referenced distribution is the matrix von Mises Fisher distribution (sometimes called the Langevin distribution) introduced by Downs (1972). Likelihood and other types of inference for this model have been considered by Khatri and Mardia

(1977), Jupp and Mardia (1979), Prentice (1986), Jupp and Mardia (1979), Chang and Bingham (1996, i.e., Bayes), Rancourt et al. (2000), Chang and Rivest (2001, i.e., M-estimation), Chikuse (2003) and Camano-Garcia (2006). Other models developed for orientation data include the Cayley distribution (León et al., 2006) and, from the materials science literature, Bunge's Gaussian and isotropic Gaussian distributions (cf. Bunge, 1982; Matthies et al., 1988; Savyolova and Nikolayev, 1995). See (Mardia and Jupp, 2000, sec. 13.2.1) for a statistical introduction to the topic of rotation matrices and Qiu et al. (2014) for a summary of one-sample Bayes inference for UARS models.

2.2 Models for Equivalence Classes of Rotation Matrices

2.2.1 UARS$([\boldsymbol{S}], \kappa)$ Models for Equivalence Classes of Rotation Matrices

Let Ω be the collection of all 3×3 rotation matrices, i.e., $SO(3)$. For a rotation matrix $\boldsymbol{O} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \Omega$ where $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are the 3 columns of $\boldsymbol{O}$, we define the equivalence class $[\boldsymbol{O}]$ of $\boldsymbol{O}$ as

$$
\begin{aligned}
[\boldsymbol{O}] = \{&(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), (\boldsymbol{y}, -\boldsymbol{x}, \boldsymbol{z}), (-\boldsymbol{x}, -\boldsymbol{y}, \boldsymbol{z}), (-\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}), (-\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x}), (\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}), \\
&(\boldsymbol{z}, -\boldsymbol{y}, \boldsymbol{x}), (-\boldsymbol{y}, -\boldsymbol{z}, \boldsymbol{x}), (\boldsymbol{x}, -\boldsymbol{z}, \boldsymbol{y}), (-\boldsymbol{z}, -\boldsymbol{x}, \boldsymbol{y}), (-\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}), (\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{y}), \\
&(\boldsymbol{z}, \boldsymbol{y}, -\boldsymbol{x}), (\boldsymbol{y}, -\boldsymbol{z}, -\boldsymbol{x}), (-\boldsymbol{z}, -\boldsymbol{y}, -\boldsymbol{x}), (-\boldsymbol{y}, \boldsymbol{z}, -\boldsymbol{x}), (\boldsymbol{x}, \boldsymbol{z}, -\boldsymbol{y}), (\boldsymbol{z}, -\boldsymbol{x}, -\boldsymbol{y}), \\
&(-\boldsymbol{x}, -\boldsymbol{z}, -\boldsymbol{y}), (-\boldsymbol{z}, \boldsymbol{x}, -\boldsymbol{y}), (\boldsymbol{x}, -\boldsymbol{y}, -\boldsymbol{z}), (-\boldsymbol{y}, -\boldsymbol{x}, -\boldsymbol{z}), (-\boldsymbol{x}, \boldsymbol{y}, -\boldsymbol{z}), (\boldsymbol{y}, \boldsymbol{x}, -\boldsymbol{z})\}.
\end{aligned} \tag{2.1}
$$

This represents the set of 24 crystallographically-symmetric rotations which are indistinguishable from $\boldsymbol{O}$ when the sides or faces of a cube are indistinguishable. Let $[\Omega]$ be the set containing all such equivalence classes of rotations, i.e. $[\Omega] = \{[\boldsymbol{O}] : \boldsymbol{O} \in \Omega\}$. We wish to develop probability models for observations $[\boldsymbol{O}]$ in $[\Omega]$ (i.e., data as unlabeled orientations) which have a central direction parameter $[\boldsymbol{S}]$ in $[\Omega]$ and a concentration parameter $\kappa > 0$ controlling spread (with concentration increasing in κ). To this end, for given parameters $[\boldsymbol{S}] \in [\Omega]$ and $\kappa > 0$, we can induce a uniform-axis-random-spin distribution on $[\Omega]$ (denoted by UARS$([\boldsymbol{S}], \kappa)$) as the distribution of $[\boldsymbol{O}]$ where $\boldsymbol{O} \sim \text{UARS}(\boldsymbol{T}, \kappa)$ and $\boldsymbol{T} \in [\boldsymbol{S}]$; see Appendix A for details on the UARS model for rotations $\boldsymbol{O}$. That is, we essentially suppose that an underlying unobservable

random rotation $\boldsymbol{O}$ is generated by a UARS model, but the observable quantity is $[\boldsymbol{O}]$, consisting of $\boldsymbol{O}$ and all other rotations in the equivalence class (2.1) of $\boldsymbol{O}$. (This formulation is similar to creating a class of distributions on $(0,\infty)$ for a random variable Y, by defining $Y = |X|$ for a normal X with mean $\mu(\neq 0)$ and a standard deviation σ; note that there are two location values $\pm|\mu|$ associated with any distribution on $(0,\infty)$ from this class, which is the analog of the location parameter $[\boldsymbol{S}]$ for unlabeled orientations $[\boldsymbol{O}] \sim UARS([\boldsymbol{S}],\kappa)$, while κ is an analog for σ^{-1}.) This distribution of $[\boldsymbol{O}]$ has a density with respect to the uniform distribution on $[\Omega]$ of the form

$$g([\boldsymbol{O}]|[\boldsymbol{S}],\kappa) = \frac{1}{24}\sum_{j=1}^{24}\frac{4\pi}{3-\mathrm{tr}(\boldsymbol{T}^T\boldsymbol{O}^{(j)})}C\left(\arccos(\frac{\mathrm{tr}(\boldsymbol{T}^T\boldsymbol{O}^{(j)})-1}{2})|\kappa\right), \tag{2.2}$$

where $C(\cdot|\kappa)$ is a circular distribution density (with respect to the Lebesgue measure) on $(-\pi,\pi]$, $\boldsymbol{O}^{(j)}$ runs through the 24 elements of $[\boldsymbol{O}]$, and $\boldsymbol{T}$ is any element of $[\boldsymbol{S}]$. (We note that expression (2.2) is reminiscent of the form of a mixture distribution on orientations but *is not*. It is a density for equivalence class of orientations following from the rule of total probability, just as a density for $Y = |X|$ above would be a sum of two normal densities.)

For purposes of exposition in the following, we focus on one particular choice of UARS model for equivalence classes of rotations, the symmetric matrix Fisher von Mises distribution on $[\Omega]$ (denoted by SMF$([\boldsymbol{S}],\kappa)$) induced from the symmetric Fisher von Mises matrix distribution on Ω (denoted by SMF$(\boldsymbol{S},\kappa)$). The matrix Fisher distribution, introduced by Downs (1972), is the oldest and arguably most popular model for orientation data in the statistics literature (cf. Chikuse, 2003), and its symmetric form has a density for rotations (see Appendix A) given by

$$f(\boldsymbol{O}|\boldsymbol{S},\kappa) = \frac{\exp\left(\kappa(\mathrm{tr}(\boldsymbol{S}^T\boldsymbol{O})-1)\right)}{I_0(2\kappa)-I_1(2\kappa)},\ \boldsymbol{O},\boldsymbol{S}\in\Omega,\ \kappa>0,$$

with a corresponding Lebesgue density for r in the UARS construction given by

$$C(r|\kappa) = \frac{(1-\cos r)\exp(2\kappa\cos r)}{2\pi(I_0(2\kappa)-I_1(2\kappa))},\ r\in(-\pi,\pi],\ \kappa>0, \tag{2.3}$$

where $I_0(z)$ and $I_1(z)$ denote the modified Bessel functions of order 0 and 1, respectively. The variance of $C(r|\kappa)$ decreases as the concentration parameter κ increases as shown in Figure 2.2. The symmetric matrix Fisher von Mises distribution on $[\Omega]$ has density (2.2) given by

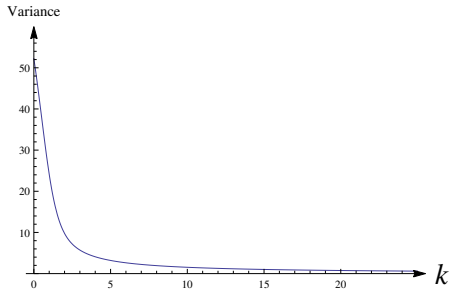

Figure 2.2 Variance of $C(r|\kappa)$ against κ.

$$g([\boldsymbol{O}]|[\boldsymbol{S}],\kappa) = \frac{1}{24}\sum_{j=1}^{24} \frac{\exp\left(\kappa\left(\mathrm{tr}(\boldsymbol{S}^T O^{(j)}) - 1\right)\right)}{I_0(2\kappa) - I_1(2\kappa)},\ [\boldsymbol{O}] \in [\Omega], \tag{2.4}$$

where $\boldsymbol{O}^{(j)}$ runs through $[\boldsymbol{O}]$ and $\boldsymbol{S}$ is any element of $[\boldsymbol{S}]$. A referee has kindly noted that (2.4) resembles a crystal symmetric standard distribution as propagated by Matthies et al. (1987). Note as well that while the Fisher von Mises model is popular Bingham et al. (2009a) demonstrate that it is not always adequate for modeling crystal orientations and the general form with other choices of $C(\cdot|\kappa)$ can be essential.

2.2.2 Likelihood-based Confidence Regions for the SMF$([\boldsymbol{S}],\kappa)$ Distribution on $[\Omega]$

We here briefly consider large-sample methods based on maximum likelihood estimation to provide a basis of comparison for a Bayes inference method to be introduced later. Computations with maximum likelihood typically involve reparameterization of the rotation matrices that are the central direction parameters in order to do numerical optimization. We use the Euler angle representation ("zxz" convention) because it is common in texture analysis in materials science,

$$[\boldsymbol{S}] = [S(\alpha,\beta,\gamma)] = \left[\begin{pmatrix} c_\alpha c_\gamma - s_\alpha s_\gamma c_\beta & s_\alpha c_\gamma + c_\alpha s_\gamma c_\beta & s_\gamma s_\beta \\ -c_\alpha s_\gamma - s_\alpha c_\gamma c_\beta & -s_\alpha s_\gamma + c_\alpha c_\gamma c_\beta & c_\gamma s_\beta \\ s_\alpha s_\beta & -c_\alpha s_\beta & c_\beta \end{pmatrix}\right], \tag{2.5}$$

where $s_\theta = \sin\theta$ and $c_\theta = \cos\theta$ for $\theta \in \{\alpha, \beta, \gamma\}$. For i.i.d. observed equivalence classes $[\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n]$ from SMF$([\boldsymbol{S}], \kappa)$ on $[\Omega]$, the log likelihood function from (2.4) is

$$l(\kappa, [\boldsymbol{S}]) = -n \log\left(b(\kappa)\right) + \sum_{i=1}^{n} \log\left(\sum_{j=1}^{24} \exp\left\{\kappa \operatorname{tr}(\boldsymbol{S}^T \boldsymbol{O}_i^{(j)})\right\}\right) + c, \tag{2.6}$$

where

$$b(\kappa) = e^{\kappa}\left(I_0(2\kappa) - I_1(2\kappa)\right), \tag{2.7}$$

$[\boldsymbol{S}]$ is expressed in Euler angle form as in (2.5), c is a constant that does not depend on κ and $[\boldsymbol{S}]$, and $\boldsymbol{S}$ and $\boldsymbol{O}^{(j)}$ follow previous conventions.

Let $\boldsymbol{\theta} = (\kappa, \alpha, \beta, \gamma)^T$, and $\hat{\boldsymbol{\theta}} = (\hat\kappa, \hat\alpha, \hat\beta, \hat\gamma)^T$ be a maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$. The nominally $1-\delta$ level large sample likelihood ratio-based confidence regions we use for κ and $[\boldsymbol{S}]$ are

$$\left\{\kappa : -2\left(l(\kappa, \hat\alpha(\kappa), \hat\beta(\kappa), \hat\gamma(\kappa)) - l(\hat\kappa, \hat\alpha, \hat\beta, \hat\gamma)\right) \le \chi^2_1(1-\delta)\right\}$$

and

$$\left\{[\boldsymbol{S}] : -2\left(l(\hat\kappa(\alpha,\beta,\gamma), \alpha, \beta, \gamma) - l(\hat\kappa, \hat\alpha, \hat\beta, \hat\gamma)\right) \le \chi^2_3(1-\delta) \text{ for some } \boldsymbol{S}(\alpha,\beta,\gamma) \in [\boldsymbol{S}]\right\} \tag{2.8}$$

respectively, where $\hat\alpha(\kappa)$, $\hat\beta(\kappa)$ and $\hat\gamma(\kappa)$ are (some) restricted MLE's of α, β and γ respectively when κ is fixed, and $\hat\kappa(\alpha,\beta,\gamma)$ is the restricted MLE of κ when α, β and γ are fixed. Using (2.6) and the observed inverse Fisher information matrix

$$\begin{pmatrix} \hat\sigma^2_{11}(\hat{\boldsymbol{\theta}}) & \widehat\Sigma_{12}(\hat{\boldsymbol{\theta}}) \\ \widehat\Sigma_{21}(\hat{\boldsymbol{\theta}}) & \widehat\Sigma_{22}(\hat{\boldsymbol{\theta}}) \end{pmatrix} = \left(-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\Big|_{\theta=\hat{\boldsymbol{\theta}}}\right)^{-1} = -\left(\begin{bmatrix} \frac{\partial^2 l}{\partial\kappa\partial\kappa} & \frac{\partial^2 l}{\partial\kappa\partial\alpha} & \frac{\partial^2 l}{\partial\kappa\partial\beta} & \frac{\partial^2 l}{\partial\kappa\partial\gamma} \\ \frac{\partial^2 l}{\partial\alpha\partial\kappa} & \frac{\partial^2 l}{\partial\alpha\partial\alpha} & \frac{\partial^2 l}{\partial\alpha\partial\beta} & \frac{\partial^2 l}{\partial\alpha\partial\gamma} \\ \frac{\partial^2 l}{\partial\beta\partial\kappa} & \frac{\partial^2 l}{\partial\beta\partial\alpha} & \frac{\partial^2 l}{\partial\beta\partial\beta} & \frac{\partial^2 l}{\partial\beta\partial\gamma} \\ \frac{\partial^2 l}{\partial\gamma\partial\kappa} & \frac{\partial^2 l}{\partial\gamma\partial\alpha} & \frac{\partial^2 l}{\partial\gamma\partial\beta} & \frac{\partial^2 l}{\partial\gamma\partial\gamma} \end{bmatrix}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^{-1},$$

the nominally $1-\delta$ level Wald large sample confidence regions that we use for κ and $[\boldsymbol{S}]$ are given by

$$\left\{\kappa : \hat\kappa - z_{1-\delta/2}\hat\sigma_{11}(\hat{\boldsymbol{\theta}}) \le \kappa \le \hat\kappa + z_{1-\delta/2}\hat\sigma_{11}(\hat{\boldsymbol{\theta}}) \text{ for some MLE } \hat{\boldsymbol{\theta}} = (\hat\kappa, \hat\alpha, \hat\beta, \hat\gamma)^T\right\}$$

and

$$\left\{[\boldsymbol{S}] : (\hat\alpha - \alpha, \hat\beta - \beta, \hat\gamma - \gamma)\widehat\Sigma^{-1}_{22}(\hat{\boldsymbol{\theta}})(\hat\alpha - \alpha, \hat\beta - \beta, \hat\gamma - \gamma)^T \le \chi^2_3(1-\delta) \right.$$
$$\left. \text{for some MLE } \hat{\boldsymbol{\theta}} = (\hat\kappa, \hat\alpha, \hat\beta, \hat\gamma)^T \text{ and some } S(\alpha,\beta,\gamma) \in [\boldsymbol{S}]\right\} \tag{2.9}$$

respectively. This model is regular, and one can expect with some effort to be able to establish the asymptotic validity of the likelihood-based confidence regions used here. But we use these regions for comparison purposes only, we do not include any such rigorous justification.

2.3 One-Sample Bayes Inference for the SMF$([\boldsymbol{S}], \kappa)$ Distribution on $[\Omega]$

2.3.1 Cone-Shaped Credible Regions for $[\boldsymbol{S}]$

Before describing the full Bayes method and posterior sampling procedure, we mention that the equivalence class $[\boldsymbol{S}]$ denoting a location parameter is a geometrically unusual model parameter. One might consider making credible regions for $[\boldsymbol{S}]$ based directly on some (highest posterior density or other) general principle. However, the marginal highest posterior region for $[\boldsymbol{S}]$ requires us to integrate out κ numerically, which is computationally intractable. Also practical visualization and interpretation of the resulting regions is geometrically challenging. For purposes of producing easily interpreted regions for $[\boldsymbol{S}]$, we adapt the concept of cone-shaped credible regions used in Bingham et al. (2009b). For $[\boldsymbol{S}], [\boldsymbol{T}] \in [\Omega]$, we define a measure of closeness of (the minmax angle between) $[\boldsymbol{S}]$ and $[\boldsymbol{T}]$ as

$$\mathrm{MMA}([\boldsymbol{S}], [\boldsymbol{T}]) = \min_{\boldsymbol{S} \in [\boldsymbol{S}], \boldsymbol{T} \in [\boldsymbol{T}]} \mathrm{MA}(\boldsymbol{S}, \boldsymbol{T}), \tag{2.10}$$

where $\mathrm{MA}(\boldsymbol{S}, \boldsymbol{T})$ is the maximum of absolute values of the 3 angles between each of the 3 pairs of perpendicular axes represented by the columns of $\boldsymbol{S}$ and $\boldsymbol{T}$. That is, if $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$ and $\boldsymbol{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3)$, let $\theta_i \in (-\pi, \pi]$ be the angle between vectors $\boldsymbol{s}_i$ and $\boldsymbol{t}_i$, $i = 1, 2, 3$,

$$\mathrm{MA}(\boldsymbol{S}, \boldsymbol{T}) = \max\{|\theta_1|, |\theta_2|, |\theta_3|\}. \tag{2.11}$$

The smaller the minmax angle between two rotation matrix classes, the closer together they are. (For the reader's reference, we note that it is possible to show that the greatest possible minmax angle is $60^\circ = 1.047$ rad.)

For $\widehat{[\boldsymbol{S}]}$ a posterior estimate of $[\boldsymbol{S}]$, we construct a $1 - \delta$ level credible region for $[\boldsymbol{S}]$ as

$$C_\rho = \{[\boldsymbol{T}] \in [\Omega] : \mathrm{MMA}([\boldsymbol{T}], \widehat{[\boldsymbol{S}]}) < \rho\}$$

where ρ is chosen so that the posterior probability of C_ρ is $1 - \delta$ (ρ is the upper δ point of the posterior distribution of $\mathrm{MMA}([\boldsymbol{T}], \widehat{[\boldsymbol{S}]})$). Figure 2.3 shows what the credible region "looks"

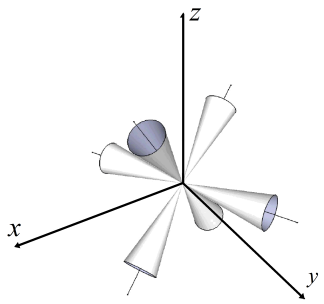

Figure 2.3 Cone-shaped region for $[\boldsymbol{S}]$.

like. The cones in Figure 2.3 center on the legs of a child's "jack" representing $\widehat{[\boldsymbol{S}]}$. Any rotation equivalence class with columns in one of its matrix representations that are contained within three cones of the "jack" lies inside the credible region.

We originally also wanted to construct cone-shaped confidence regions for $[\boldsymbol{S}]$ containing likelihood-based regions (2.8) and (2.9). However, their identification was computationally infeasible and too expensive for the method based on likelihood ratio tests, so we consider here only the raw confidence regions (2.8) and (2.9) for $[\boldsymbol{S}]$ as methods based on the likelihood. (These do not have similarly obvious geometric interpretations.)

2.3.2 One-Sample Bayes Methods

First we want to choose a reasonable (and possibly non-informative) prior. It is natural and consistent with the methods of Bingham et al. (2009b) to make the priors for κ and $\boldsymbol{S}$ independent (a prior for $\boldsymbol{S}$ induces a corresponding prior for $[\boldsymbol{S}]$). For $\boldsymbol{S}$, a natural choice is the uniform distribution on Ω; as densities on Ω are formulated with the uniform distribution as the dominating measure (as in (A.2)and (2.2), the uniform distribution as a density 1 with respect to itself. We considered deriving a Jeffreys prior for κ from (2.4), but

$$\mathrm{E}\left(\frac{\partial^2}{\partial\kappa^2}\log\left(g([\boldsymbol{O}],[\boldsymbol{S}]\mid\kappa)\right)\right)=\frac{\partial^2}{\partial\kappa^2}\log\left(b(\kappa)\right)+\mathrm{E}\left(\frac{\partial^2}{\partial\kappa^2}\log\left(\sum_{j=1}^{24}\exp\left\{\kappa\,\mathrm{tr}(\boldsymbol{S}^T\boldsymbol{O}^{(j)})\right\}\right)\right) \tag{2.12}$$

is complicated, which suggests that the Jeffreys prior for κ is impractical (here $b(\kappa)$ is as in (2.7)). Notice that

$$\pi(\kappa) = \sqrt{-\frac{\partial^2}{\partial \kappa^2} \log\left(b(\kappa)\right)}$$

is the Jeffreys prior for κ for the Cir(κ) distribution from (2.3) on $(-\pi, \pi]$. So we simply drop the second term in (2.12) and use $\pi(\kappa)$ as a prior for κ and treat $[\boldsymbol{S}]$ and κ as *a priori* independent.

For the chosen prior combination for $(\kappa, [\boldsymbol{S}])$, the posterior density can be written as

$$p(\kappa, [\boldsymbol{S}]) \propto L(\kappa, [\boldsymbol{S}])\pi(\kappa),$$

where $L(\kappa, [\boldsymbol{S}]) = \exp\{l(\kappa, [\boldsymbol{S}])\}$ is the likelihood for κ and $[\boldsymbol{S}]$ from (2.6). It can be shown that the posterior density $p(\kappa, [\boldsymbol{S}])$ is proper (see the Appendix of Bingham et al., 2009b for the proof of a similar result for orientations). To get draws from the posterior distribution, we can use a Metropolis-Hastings-within-Gibbs algorithm:

1. Set starting values (denoted by $[\boldsymbol{S}]^0$ and κ^0) for $[\boldsymbol{S}]$ and κ.
2. Generate $[\boldsymbol{S}]^{j*} \sim \text{SMF}([\boldsymbol{S}]^{j-1}, \rho)$ as a proposal for $[\boldsymbol{S}]^j$.
3. Compute $r_j^1 = \frac{p([\boldsymbol{S}]^{j*}, \kappa^{j-1})}{p([\boldsymbol{S}]^{j-1}, \kappa^{j-1})}$. If $r_j^1 >= 1$, accept $[\boldsymbol{S}]^{j*}$ as $[\boldsymbol{S}]^j$; otherwise, accept $[\boldsymbol{S}]^{j*}$ with probability r_j^1.
4. Generate $\kappa^{j*} \sim \text{lognormal}\left(\log(\kappa^{j-1}), \sigma^2\right)$ as a proposal for κ^j.
5. Compute $r_j^2 = \frac{p(S^j, \kappa^{j*})\kappa^{j*}}{p(s^j, \kappa^{j-1})\kappa^{j-1}}$. If $r_j^2 >= 1$, accept κ^{j*} as κ^j; otherwise, accept κ^{j*} with probability r_j^2.

The posterior draw yielding the largest joint posterior density is an approximation to the posterior mode estimate. This approximate posterior mode estimate for $[\boldsymbol{S}]$ is used to construct a cone-shaped credible region for $[\boldsymbol{S}]$ as described in Chapter 2.3.1, and a one-sided (lower limit) credible interval is used for κ in this manuscript.

2.3.3 Simulation Results and Comparison to Likelihood-based Methods

Our major purpose in this section is to compare the Bayes method with large-sample maximum likelihood-based methods. The two different likelihood-based confidence regions are constructed as in (2.8) and (2.9), and the Bayes credible regions are constructed as described in Chapters 2.3.1 and 2.3.2. For several combinations of choices of κ and sample size n, we simulated 1,000 samples from the SMF($[\boldsymbol{S}], \kappa$) distribution on $[\Omega]$ as described in Chapter 2.3.2. We held

$$[\boldsymbol{S}] = [S(2.3, 1.1, 5.9)]$$

in (2.5) for convenience in comparing our results with the results in Bingham et al. (2009b). For each sample generated from SMF($[\boldsymbol{S}], \kappa$), we simulated 20,000 draws from the posterior distribution. All of our inferences were based on these posterior draws after a burn-in of 5,000 iterations (again these values were chosen for comparison of our results to those in Bingham et al., 2009b). We used true values as starting values for $[\boldsymbol{S}]$ and κ. (We note this for completeness. After burn-in this start has no effect on our results.) Tuning parameters (see Table 2.1) were chosen to make the jumping rates of Metropolis-Hastings steps to be around 0.4 (see Gelman et al., 2004, sec. 11.10). For $\kappa \geq 5$ and $\kappa \geq 4$ with sample size $n = 100$ the tuning parameters in Table 2.1 work well. For other cases, the tuning parameters in Table 2.1 were the (baseline) tuning parameters we used in the simulation, but they did not produce jumping rates around 0.4 for every simulated data set. If tuning parameters had been chosen separately for each simulated data set in these cases, the performance of the Bayes method would presumably have looked better (in terms of producing coverage rates closer to the nominal rate).

We computed the coverage rates for κ and $[\boldsymbol{S}]$ for each combination of κ and n based on the 1,000 simulation runs (see Table 2.2). From the simulation results summarized in Table 2.2, we can see that the performances of these three methods all depend on the sizes of κ and n. Generally speaking, these methods work for $\kappa \geq 4$, and for $\kappa \geq 3$ with sample size $n = 100$. For κ small, the samples tend to be very spread out and none of the three methods works particulary well for small κ (in terms of having actual coverage rates matching the nominal rate). An interesting phenomenon is that while the likelihood-based confidence regions under-

cover the true parameters, the Bayes credible regions tend to over-cover the true parameters in these cases. For a fixed relative large κ (say, $\kappa \geq 4$), all three methods work better for larger sample sizes, which is intuitively plausible. However, as n decreases, the performances of the likelihood-based methods degrade substantially. This is easy to understand, as the two likelihood-based methods are large-sample methods. Although the coverage rate for the Bayes method also worsens as n decreases, the deterioration is relatively mild. Usually the Bayes credible regions slightly over-cover the true parameters as the sample size, n, decreases. Overall, the Bayes method seems to beat the likelihood-based methods in terms of achieving a nominal coverage level.

	$n = 10$		$n = 30$		$n = 100$	
	ρ	σ	ρ	σ	ρ	σ
$\kappa = 1$	1e-10	2.5	1e-10	2.5	1e-10	2.5
$\kappa = 3$	1e-10	2.5	1e-10	1.5	25	0.5
$\kappa = 4$	1e-8	2	12	0.8	85	0.35
$\kappa = 5$	9.65	1.5	48	0.7	115	0.4
$\kappa = 20$	100	0.85	400	0.5	1,000	0.28

Table 2.1 Values of tuning parameters ρ and σ.

2.4 Comparison to Inferences Based on SMF($\boldsymbol{S}, \kappa$) Models on Ω for Preprocessed Data

Intuitively, for large concentration parameters κ (i.e., small variability in equivalence classes) we should get inferences about κ based on an i.i.d. SMF($[\boldsymbol{S}], \kappa$) model for observations (i.e., unlabeled orientations) that are similar to those based on treating preprocessed versions of representatives of equivalence classes as i.i.d. SMF($\boldsymbol{S}, \kappa$) observations (i.e., orientations). That is, when κ is large, we do not expect to see any real impact of *ad hoc* treatment of the labeling issue. However, when κ is small and the model is SMF($[\boldsymbol{S}], \kappa$), inference for κ based on treating representatives of equivalence classes of rotations as if they were generated from the SMF($\boldsymbol{S}, \kappa$) model will potentially be ineffective even if we preprocess data as described in the introduction. We illustrate these points with comparisons based on real and simulated data sets in the

(n, κ)	Bayes		LRT		Wald	
	κ	$\boldsymbol{S}$	κ	$\boldsymbol{S}$	κ	$\boldsymbol{S}$
(10,1)	98.8	91.0	35.2	82.6	17.8	43.2
(30,1)	99.4	92.1	39.4	87.0	11.6	53.2
(100,1)	99.9	94.1	43.1	84.7	12.6	59.1
(10,3)	99.0	97.7	85.4	87.9	93.9	70.2
(30,3)	97.3	98.3	92.7	93.0	94.9	87.6
(100,3)	96.8	96.8	94.8	94.5	95.5	96.1
(10,4)	96.9	98.9	90.2	90.7	95.8	81.1
(30,4)	96.1	96.1	94.6	93.3	93.7	93.7
(100,4)	95.9	95.0	95.7	93.8	95.7	95.1
(10,5)	95.0	97.5	90.8	91.4	96.4	85.5
(30,5)	95.3	95.1	93.6	94.5	95.6	95.0
(100,5)	95.6	95.4	94.9	95.2	95.3	97.0
(10,20)	94.0	94.7	89.9	93.7	94.0	92.3
(30,20)	94.6	94.5	93.9	94.3	96.1	94.8
(100,20)	95.2	95.0	94.7	95.2	95.4	95.8

Table 2.2 Coverage rates (as percentages) for κ and $[\boldsymbol{S}]$ for nominally 95% Bayes regions, inverted likelihood ratio test regions and Wald regions, for some choices of (n, κ).

following subsections.

2.4.1 A First Comparison Based on Repeated Measurements

We applied Bayes methods based on both the SMF$(\boldsymbol{S}, \kappa)$ and SMF$([\boldsymbol{S}], \kappa)$ models to the preprocessed EBSD data given in Table 1 of Bingham et al. (2009a). These are 14 (preprocessed) observations representing "repeat" measurements at a single location (they come from 14 different scans of the specimen). The sample of orientations is highly concentrated, so we expect the Bayes methods based on SMF$(\boldsymbol{S}, \kappa)$ and SMF$([\boldsymbol{S}], \kappa)$ models to give similar inferences for κ. We simulated 20,000 posterior draws for both models. The histograms for κ (after a burn-in of 5,000) for the two analyses are presented in Figure 2.4. The histograms are similar. The posterior mean estimates of κ are $\hat{\kappa} = 120,493$ for the SMF$(\boldsymbol{S}, \kappa)$ analysis and $\hat{\kappa} = 119,735$ for the SMF$([\boldsymbol{S}], \kappa])$ analysis, which are also similar. The posterior mode estimate of $\boldsymbol{S}$ for the

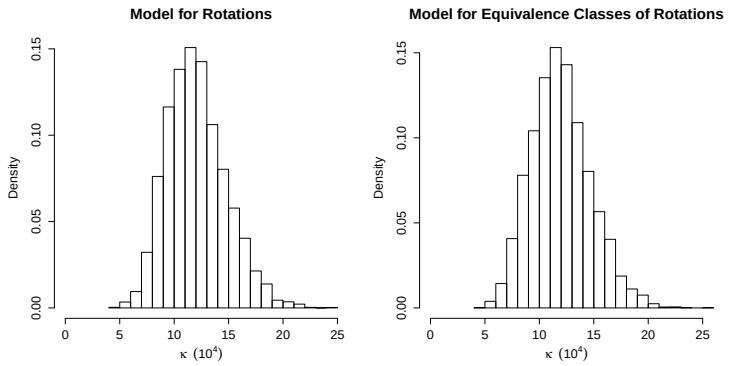

Figure 2.4 Histograms of posterior draws for κ for SMF($\boldsymbol{S}, \kappa$) and SMF($[\boldsymbol{S}], \kappa$) models (repeated real measurements example).

SMF($\boldsymbol{S}, \kappa$) model (or equivalently the MLE of $\boldsymbol{S}$ under this model) is

$$\hat{\boldsymbol{S}}_1 = \begin{pmatrix} -0.647 & -0.282 & -0.708 \\ 0.688 & -0.617 & -0.383 \\ -0.329 & -0.734 & 0.594 \end{pmatrix},$$

and the posterior mode estimate of $[\boldsymbol{S}]$ for the SMF($[\boldsymbol{S}], \kappa$) model is

$$\widehat{[\boldsymbol{S}_2]} = \left[\begin{pmatrix} -0.647 & -0.282 & -0.708 \\ 0.687 & -0.617 & -0.383 \\ -0.329 & -0.734 & 0.594 \end{pmatrix} \right].$$

The estimate of $\boldsymbol{S}$ for the SMF($\boldsymbol{S}, \kappa$) model here is not meaningful in practice unless we treat it as representating an equivalence class of rotations, because if a different reference rotation matrix had been chosen in the preprocessing, the estimate of $\boldsymbol{S}$ for the SMF($\boldsymbol{S}, \kappa$) model might have been different here. The minmax angle (as defined in Chapter 2.3.1) between $[\hat{\boldsymbol{S}}_1]$ and $\widehat{[\boldsymbol{S}_2]}$ is $\text{MMA}([\hat{\boldsymbol{S}}_1], \widehat{[\boldsymbol{S}_2]}) = 1.17 \times 10^{-4}$ rad, which is very small. It was suggested in Bingham et al. (2009a) that a (non-regular) von Mises UARS($\boldsymbol{S}, \kappa$) model fits these EBSD data better than the SMF($\boldsymbol{S}, \kappa$) model. We have verified that the Bayes inferences based on UARS($\boldsymbol{S}, \kappa$) and UARS($[\boldsymbol{S}], \kappa$) models are also similar for this other (von Mises) choice of circular distribution defining the UARS class.

2.4.2 Another Comparison Based on Real Data

We chose a large 10×10 lattice out of the 3,449 locations measured on the surface of the nickel specimen. Using the first scan at each location, we have 100 observations of equivalence classes of rotations. We applied the two different Bayes methods of Chapter 2.4.1 to this data set. We simulated 20,000 posterior draws for both methods. The histograms for κ (after a burn-in of 5,000) for the two analyses are presented in Figure 2.5. The posterior mean estimate of κ is $\hat{\kappa} = 4.696$ for the SMF$(\boldsymbol{S}, \kappa)$ model and $\hat{\kappa} = 3.807$ for the SMF$([\boldsymbol{S}], \kappa)$ model. The 95% two-sided credible interval for κ based on the SMF$([\boldsymbol{S}], \kappa)$ model is $(3.062, 4.635)$. We can see that the estimates of concentration parameter κ based on the two different models are significantly different. The posterior mode estimate of $\boldsymbol{S}$ for the SMF$(\boldsymbol{S}, \kappa)$ model (or equivalently the MLE of $\boldsymbol{S}$ under this model) is

$$\hat{\boldsymbol{S}}_1 = \begin{pmatrix} -0.830 & -0.034 & -0.557 \\ 0.427 & -0.681 & -0.595 \\ -0.360 & -0.731 & 0.580 \end{pmatrix},$$

and the posterior mode estimate of $[\boldsymbol{S}]$ for the SMF$([\boldsymbol{S}], \kappa)$ model is

$$\widehat{[\boldsymbol{S}]}_2 = \begin{pmatrix} -0.317 & 0.314 & 0.895 \\ -0.806 & -0.586 & -0.080 \\ 0.499 & -0.747 & 0.439 \end{pmatrix}.$$

The 95% cone-shaped credible region for $\boldsymbol{S}$ based on the SMF$([\boldsymbol{S}], \kappa)$ model has a minmax angle cut off at 0.3377 rad $= 19.35°$. As in Chapter 2.4.1, the estimate of $\boldsymbol{S}$ for the SMF$(\boldsymbol{S}, \kappa)$ model is not meaningful in practical terms, unless we treat it as representing an equivalence class of rotations. The minmax angles between $[\hat{\boldsymbol{S}}_1]$ and $\widehat{[\boldsymbol{S}_2]}$ is MMA$([\hat{\boldsymbol{S}}_1], \widehat{[\boldsymbol{S}_2]}) = 0.3633$ rad $= 20.82\,°$, which indicates a significant difference between the two estimates. This example shows that inferences based on these two different Bayes methods can be much different for real data that are not highly concentrated.

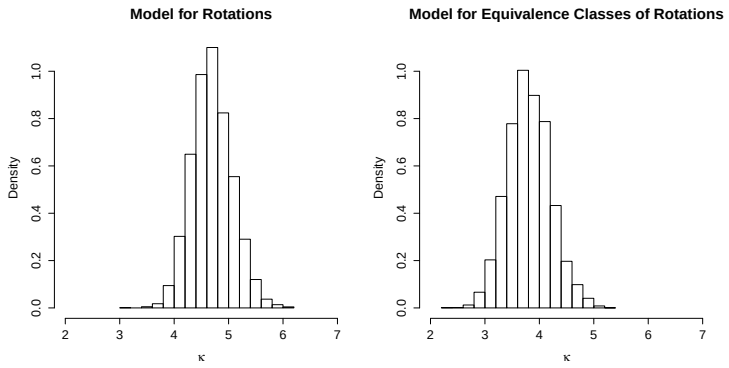

Figure 2.5 Histograms of posterior draws for κ for SMF($\boldsymbol{S}, \kappa$) and SMF($[\boldsymbol{S}], \kappa$) models (second real data example).

2.4.3 A Comparison Based on a Small Simulation

In Chapter 2.3.3, we conducted simulations for various choices of concentration parameter κ and sample size n. To further establish (on the basis of a case where the truth is known) that it is not always effective to apply an analysis based on the SMF($\boldsymbol{S}, \kappa$) model for preprocessed equivalence classes of rotations from the SMF($[\boldsymbol{S}], \kappa$) model, we applied the Bayes method based on the SMF($\boldsymbol{S}, \kappa$) model to a preprocessed (all rotation matrices were lined up with the first observation as mentioned in Chapter 2.1) representative rotation matrix (equivalence class) sample with $\kappa = 5$ and sample size $n = 100$ from the simulation in Chapter 2.3.3. We also simulated 20,000 posterior draws for the Bayes method based on the SMF($\boldsymbol{S}, \kappa$) model. The histograms for κ (after a burn-in of 5,000) for the two models are shown in Figure 2.6. The posterior mean estimate of κ is $\hat{\kappa} = 3.3933$ based on the SMF($\boldsymbol{S}, \kappa$) model and $\hat{\kappa} = 5.1568$ for the (true) SMF($[\boldsymbol{S}], \kappa$) model. The posterior mode estimates of $\boldsymbol{S}$ are

$$\hat{\boldsymbol{S}}_1 = \begin{pmatrix} 0.462 & -0.847 & 0.261 \\ -0.44e & -0.476 & -0.760 \\ 0.768 & 0.236 & -0.595 \end{pmatrix},$$

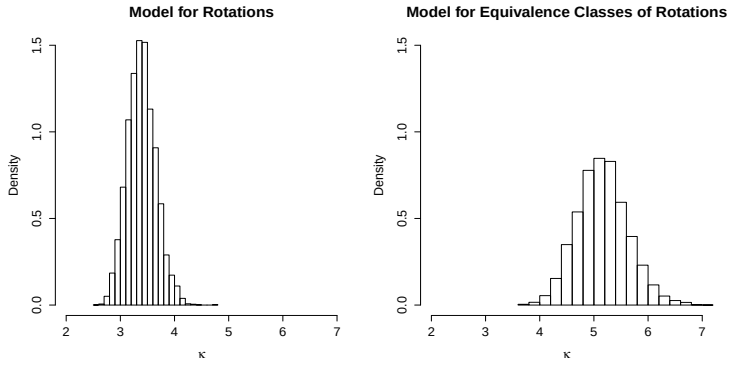

Figure 2.6 Histograms of posterior draws for κ for SMF$(\boldsymbol{S}, \kappa)$ and SMF$([\boldsymbol{S}], \kappa)$ models (simulated data example).

for the SMF$(\boldsymbol{S}, \kappa)$ model and

$$\widehat{[\boldsymbol{S}_2]} = \begin{pmatrix} -0.492 & 0.813 & -0.311 \\ -0.572 & -0.034 & 0.820 \\ 0.656 & 0.581 & 0.482 \end{pmatrix}.$$

for the SMF$([\boldsymbol{S}], \kappa)$ model. As in Chapter 2.4.1, the estimate of $\boldsymbol{S}$ for the SMF$(\boldsymbol{S}, \kappa)$ model is not meaningful in practical terms, unless we treat it as representing an equivalence class of rotations. The minmax angles between $[\hat{\boldsymbol{S}}_1]$, $\widehat{[\boldsymbol{S}_2]}$ and the truth $[\boldsymbol{S}] = [S(2.3, 1.1, 5.9)]$ are MMA$([\hat{\boldsymbol{S}}_1], [\boldsymbol{S}]) = 0.5937$ radian and MMA$(\widehat{[\boldsymbol{S}_2]}, [\boldsymbol{S}]) = 0.0367$ radian. These results show that the Bayes method based on the SMF$(\boldsymbol{S}, \kappa)$ model and preprocessing does not always work well when the actual data are unlabeled orientations and that inference is much better when directly handling the unlabeled nature of such data.

2.5 Extensions of Models to Other Equivalence Classes of Rotations

As described in Chapter 2.2.1, if we set up a right-hand coordinate system on an unlabeled cube, we get a rotation matrix class containing 24 equivalent elements by comparing the coordinate system on the cube to a pre-selected reference coordinate system. If we can distinguish some features of a rectangular solid (e.g. if we can distinguish some edges or sides of a rectangular solid from others), we can formulate models on equivalence classes of rotations different

from those studied thus far. For example, if we have a cuboid with two axes having the same length at a corner (see Figure 2.7), we can make it a convention to choose the axis with a

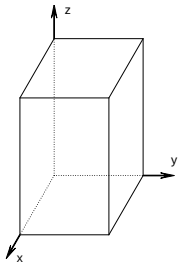

Figure 2.7 Right-hand coordinate systems on a symmetric cuboid.

length different from the others at a corner as the z-axis. Under this convention, there are 8 different ways to put a right-hand coordinate system on the cuboid, i.e. an equivalence class of rotations has 8 elements in this case. Let rotation matrix $\boldsymbol{O} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. The equivalence class $[\boldsymbol{O}]$ of $\boldsymbol{O}$ is

$$[\boldsymbol{O}] = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), (-\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}), (-\boldsymbol{x}, -\boldsymbol{y}, z), (\boldsymbol{y}, -\boldsymbol{x}, \boldsymbol{z}), \\ (\boldsymbol{y}, \boldsymbol{x}, -\boldsymbol{z}), (\boldsymbol{x}, -\boldsymbol{y}, -\boldsymbol{z}), (-\boldsymbol{y}, -\boldsymbol{x}, -\boldsymbol{z}), (-\boldsymbol{x}, y, -\boldsymbol{z})\} \tag{2.13}$$

in this case. This equivalence class corresponds to the tetragonal crystal system in crystallography. There are a number of lattice groups that fall under this crystal system. We can formulate models on equivalence classes of rotations as defined in (2.13) exactly as we did in Chapter (2.2.1). A referee has kindly further noted that in some situations, it is possible to identify the labeling of the 3 edges/axes of the coordinate system, but not their signs. There are 2 ways to choose the direction of the z-axis, so an equivalence class of rotations has 2 elements in this case. Let rotation matrix $\boldsymbol{O} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. The equivalence class $[\boldsymbol{O}]$ of $\boldsymbol{O}$ is

$$[\boldsymbol{O}] = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), (-\boldsymbol{x}, -\boldsymbol{y}, -\boldsymbol{z})\} \tag{2.14}$$

in this context. For other cases, analogous arguments are possible. In an application, we only need to identify what equivalence classes of rotations we wish to consider or allow. For

example, different crystallographic geometries may lead to different equivalence classes based on symmetries in materials science applications.

2.6 Conclusion

Models on equivalence classes of rotations provide a logically coherent way to treat data arising in some materials science applications, and their use does not require any *ad hoc* preprocessing of such data into orientations. What is more important is that for some situations such preprocessing cannot work well, and analysis based on models for equivalence classes of rotations is simply a better methodology. For the new UARS($[\boldsymbol{S}], \kappa$) models on rotation matrix equivalence classes, we have developed an effective Bayes method for one-sample inference. The Bayes method (for the SMF($[\boldsymbol{S}], \kappa$) model) generally works better than maximum likelihood-based methods (that also fail to have obvious practical interpretations), especially when the sample size is small.

CHAPTER 3. A METHOD FOR MAPPING GRAINS IN EBSD SCANS OF MATERIAL SPECIMENS USING SPATIALLY INFORMED CLUSTERING OF 3-D ORIENTATIONS

A paper to be sumbmitted to the Journal of Technometrics

Chuanlong Du [1], Daniel Nordman[2] and Stephen Vardeman [3]

Abstract

Crystals in metal specimens have internally homogeneous orientations relative to a pre-chosen reference coordinate system, and clusters of crystals in the metal with locally similar orientations constitute so-called "grains." The nature of these grains (shape, size, etc.) affects physical properties (e.g., hardness, conductivity, etc.) of the material. Electron backscatter diffraction (EBSD) machines produce measured orientations on a regular grid on a specimen surface that can be processed into a grain map, indicating different grains on the metal surface, after a material is scanned. Current algorithms for creating such maps process locations in some predetermined order and start a new grain/cluster whenever these encounter an orientation more than a fixed distance (angle) from those in a previous grain. This method is ad hoc and may not produce grain maps that accurately portray grain structure. We propose a more principled way to cluster measured orientations of locations and formulate grain maps for material specimens.

[1]Primary researcher and author, graduate student, Department of Statistics, Iowa State University.
[2]Associate Professor, Department of Statistics, Iowa State University.
[3]University Professor, Department of Statistics, Department of Industrial and Manufacturing Systems Engineering, Iowa State University.

3.1 Introduction

A common tool of material science texture analysis is the making of grain maps. A grain map indicates the positions of grains/clusters of measured orientations of crystals on the surface of a specimen, indicating groups of crystals that locally share similar orientations. EBSD is a popular way to infer orientations of crystals in material specimens. After scanning, existing EBSD machines commonly produce grain maps derived from ad hoc mapping algorithms. Although details of algorithms that existing EBSD machines use are proprietary and not completely clear to us, it appears that the algorithms typically traverse a grid of locations (in some order) and begin a new grain whenever an orientation that is more than a fixed distance (measured in terms of angle) from the previous grain is found (see Zhou and Wang, 2007, pg. 54). This method is simple and fast, but the resulting grain map might not accurately capture the underlying relationship between orientations and can produce undesirable artifacts in maps due to missing or outlying EBSD observations. When an EBSD machine scans the orientation of a crystal at a fixed location on a metal specimen, it records both the orientation information and the 2-D coordinate information. It is desirable to use both the orientation and the location information to cluster orientations on surfaces of specimens. For convenience, we create and henceforth use a terminology *orienlocation* to denote a data observation as an orientation-and-location pair. We propose a simple method that clusters orienlocations based on a combination of a penalized distance between orientations and a penalized Euclidean distance between locations.

The rest of the manuscript is organized as follows. Chapter 3.2 develops weighted distance measures for orienlocations and, based on such distances, Chapter 3.3 describes a hierarchical clustering algorithm for producing grain maps from orienlocation data. Chapter 3.4 illustrates and explains the proposed clustering method in producing grain maps for real EBSD data. Chapter 3.5 then examines the performance of the clustering method (as a function of tuning parameter choices) in a simulated data example. Concluding remarks are provided in Chapter 3.6.

3.2 Distance Between Orienlocations

In order to develop a new clustering algorithm for making grain maps out of EBSD data (orienlocations), we first need to outline some concepts of "distance," both between orientations and between 2-D locations, which can be combined into a weighted distance measure between orienlocations. Chapters 3.2.1-3.2.2 briefly describe distances on orientations from EBSD; Chapter 3.2.3 describes distances between locations. Chapters 3.2.4-3.2.5 then combine these two distance types into one weighted or penalized distance measure for orienlocations.

3.2.1 Distance Between Orientations

Orientations in 3-dimensional space are usually represented by 3×3 rotation matrices (orthogonal matrices with determinant 1). For example, the identity matrix $\boldsymbol{I}_3$ is a rotation matrix, which represents the "reference" orientation/coordinate system. If O_1 and O_2 denote two rotation matrices/orientations, a natural distance d_o between O_1 and O_2 can be defined as

$$d_o(O_1, O_2) = \sqrt{3 - tr(O_1' O_2)}, \tag{3.1}$$

where $tr(O)$ is the trace of the matrix O, i.e., the sum of diagonal elements of the matrix O. This is a scale-adjusted Frobenius distance between O_1 and O_2, which is defined as

$$d_F(O_1, O_2) = \sqrt{tr\left((O_1 - O_2)'(O_1 - O_2)\right)}. \tag{3.2}$$

The smaller is the distance between two rotation matrices/orientations, the "closer" they are. For example,

$$A = \begin{pmatrix} 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \\ -1/3 & 2/3 & 2/3 \end{pmatrix} \tag{3.3}$$

and

$$B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{3.4}$$

are two rotation matrices. From the above definition, the distance between them is

$$d_o(A, B) = \sqrt{3 - tr(A'B)} = \sqrt{3 - 2} = 1. \tag{3.5}$$

3.2.2 Distance Between Equivalence Classes of Orientations

It is argued in Du et al. (2013) that for cubic crystal structures it is best to treat orientations measured by an EBSD machine as equivalence classes of orientations and make inferences based on such classes. this is because EBSD orientation measurements are not technically true orientations (3×3 rotations matrices) but rather indicate a collection of 24 orientations (an equivalence class) which are physically equivalent representations of a crystal's orientation (due to inherent crystallographic symmetries in EBSD readings). As a consequence, it then becomes logically most coherent and practically effective to perform clustering based on equivalence classes of orientations. For an orientation represented by rotation matrix $O = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ where $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are the 3 columns of O, the equivalence class $[O]$ of O is defined in Du et al. (2013) as

$$\begin{aligned}[O] = \{&(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), (\boldsymbol{y}, -\boldsymbol{x}, \boldsymbol{z}), (-\boldsymbol{x}, -\boldsymbol{y}, \boldsymbol{z}), (-\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}), (-\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x}), (\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}),\\ &(\boldsymbol{z}, -\boldsymbol{y}, \boldsymbol{x}), (-\boldsymbol{y}, -\boldsymbol{z}, \boldsymbol{x}), (\boldsymbol{x}, -\boldsymbol{z}, \boldsymbol{y}), (-\boldsymbol{z}, -\boldsymbol{x}, \boldsymbol{y}), (-\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}), (\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{y}),\\ &(\boldsymbol{z}, \boldsymbol{y}, -\boldsymbol{x}), (\boldsymbol{y}, -\boldsymbol{z}, -\boldsymbol{x}), (-\boldsymbol{z}, -\boldsymbol{y}, -\boldsymbol{x}), (-\boldsymbol{y}, \boldsymbol{z}, -\boldsymbol{x}), (\boldsymbol{x}, \boldsymbol{z}, -\boldsymbol{y}), (\boldsymbol{z}, -\boldsymbol{x}, -\boldsymbol{y}),\\ &(-\boldsymbol{x}, -\boldsymbol{z}, -\boldsymbol{y}), (-\boldsymbol{z}, \boldsymbol{x}, -\boldsymbol{y}), (\boldsymbol{x}, -\boldsymbol{y}, -\boldsymbol{z}), (-\boldsymbol{y}, -\boldsymbol{x}, -\boldsymbol{z}), (-\boldsymbol{x}, \boldsymbol{y}, -\boldsymbol{z}), (\boldsymbol{y}, \boldsymbol{x}, -\boldsymbol{z})\}, \end{aligned} \tag{3.6}$$

corresponding to the particular orientation's 24 crystallographically equivalent representations (these are obtained by re-labeling and shuffling the coordinate axes). It is natural to define a distance $d_{[o]}$ between two equivalence classes of orientations as the minimum of distances between elements from the two equivalence classes. Mathematically, this is

$$d_{[o]}([O_1], [O_2]) = \min_{T \in [O_2]} d_o(S, T), \tag{3.7}$$

where S is any element in $[O_1]$. For neater notation, we extend the definition of $d_{[o]}$ to orientations as

$$d_{[o]}(O_1, O_2) = d_{[o]}([O_1], [O_2]). \tag{3.8}$$

Using $d_{[o]}$ as a distance between orientations, we henceforth avoid the verbosity of discussing "equivalence classes of" orientations.

3.2.3 Euclidean Distance

A natural distance between locations on a specimen surface, i.e., points in 2-dimensional space, is the Euclidean distance. For $L_1 = (x_1, y_1)$ and $L_2 = (x_2, y_2)$ two points in 2-dimensional space, the Euclidean distance d_l between them is

$$d_l(L_1, L_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{3.9}$$

For example, the Euclidean distance between point $p_1 = (1, 2)$ and $p_2 = (4, 6)$ is

$$d_l(p_1, p_2) = \sqrt{(1 - 4)^2 + (2 - 6)^2} = 5. \tag{3.10}$$

3.2.4 A Combined Distance of Orientations and Locations for Orienlocations

A rotation matrix/orientation in 3-dimensional space can be expressed by 3 independent variables, e.g., the Euler angler representation (Pio, 1966). So one might essentially treat orienlocations as observations in 5-dimensional space and define a usual Euclidean distance for them. However, a reduced-dimension representation of orientations requires one to restrict the domain of variables, which results in a space that is much different from a usual Euclidean space. Thus trying to capture "closeness" of two orienlocations, consisting of paired observations on fundamentally different manifolds in terms of a Euclidean distance in a reduced 5-dimensional space is not appealing.

A more natural approach is to define a distance for orienlocations by combining a distance for orientations and a distance for locations. Since we are *a priori* not sure how much we should rely on information from orientations relative to that for locations, it is sensible to define a weighted distance for orienlocations, and choose an appropriate weight coefficient based on real data. Let $C_1 = (O_1, L_1)$ and $C_2 = (O_2, L_2)$ be two orienlocations, where O_i is the orientation part and L_i is the location part. A weighted squared distance d_{ol} between C_1 and C_2 can be defined as

$$d^2_{ol}(C_1, C_2) = \omega\, d^2_{[o]}(O_1, O_2) + d^2_l(L_1, L_2), \tag{3.11}$$

where $\omega > 0$ is the weight coefficient. The factor ω controls the relative importances of the orientation information and location information. For a small ω, d_l plays a more important

role than $d_{[o]}$, and vice versa. We will see that in practice ω can be tuned to produce sensible grain maps in a distance-based clustering algorithm for orienlocations. For any two orientations/rotation matrices, it holds that $0 \leq d_o^2(O_1, O_2) \leq 4$. So in practice, one might also scale d_l^2 to be between 0 and 4 before beginning. This is not required, but makes it easier to choose an appropriate value for ω.

3.2.5 Penalized Distances for Orientations and Locations

In practice, one might want to control the size of grains. A good way to achieve this is to employ "penalized" versions of (squared) distances between orientations and locations. A penalized squared distance $\tilde{d}_{ol}$ between orienlocations can be defined as

$$\tilde{d}_{ol}^2(C_1, C_2) = \omega f(d_{[o]}^2(O_1, O_2), c_o, a_o) + f(d_l^2(L_1, L_2), c_l, a_l), \tag{3.12}$$

where

$$f(x, c, a) = x + a(x - c)I(x > c), \tag{3.13}$$

for constants $c_o \geq 0$, $a_o \geq 0$, $c_l \geq 0$ and $a_l \geq 0$. Factors c_i and a_i control how the squared distance d_i^2, $i \in \{[o], l\}$, is penalized. For big a_i and small c_i, the squared distance d_i^2 is penalized more, and the opposite is true for small a_i and c_i. The values a_o and a_l should typically be large so that two orientations with a squared distance greater than c_o are unlikely to be placed in a same grain/cluster, and two locations with a squared distance greater than c_l are unlikely to be placed in a same grain/cluster. When $a_i = 0$ or $c_i \geq 4$, $i \in \{[o], l\}$, the orientation/Euclidean distance is not penalized (assume the squared Euclidean distance is scaled to be between 0 and 4 as suggested in Chapter 3.2.4).

3.3 Clustering Orienlocations

With distances between each pair of orienlocations available, one can consider clustering orienlocations using various existing clustering methods in literature, e.g., the hierarchical clustering method (see Hastie et al., 2009, 14.3.12 Hierarchical Clustering) or some version of k-means algorithm (see Hastie et al., 2009, 13.2.1 K-means Clustering). Regarding the latter possibility, it is not clear to us what is a natural way to define a "center" for orientations

(represented as rotation matrices), as the simple element-wise average of rotation matrices is not necessarily a rotation matrix. (Starting with k elements of a data set as initial prototypes/centers/medoids, and restricting to only such choices in subsequent iterations of a search appears to be computationally prohibitive.) Therefore, we will here use a hierarchical clustering method. The agglomerative hierarchical clustering method is preferred in this case because it is simple and fast (compared to divisive hierarchical clustering method and a k-medoids algorithm), yields consistent results, and scales well to the need of producing grain maps with different number of grains. The agglomerative hierarchical method starts with every data point in its own grain/cluster, and then repeatedly merges the pair of grains/sets of clusters that are "closest" in some well-defined sense until all data points are in the same grain/cluster. The process of hierarchical clustering can be represented by a dendrogram, and the dendrogram can be cut at any level to produce grain maps/clusters with required numbers of clusters. To illustrate how the agglomerative hierarchical clustering method works, consider a toy example where the data points are 2-dimensional (see Table 3.1) and the Euclidean distance between data points is used. In this illustration, the complete linkage (for more on linkage, see Hastie et al., 2009, 14.3.12 Hierarchical Clustering) is used, which takes the distance between two grains/clusters G_1 and G_2 to be

$$d_g(G_1, G_2) = max\{d(x_1, x_2) : x_1 \in G_1, x_2 \in G_2\} \tag{3.14}$$

where $d(x_1, x_2)$ is the distance between data points x_1 and x_2 (the Euclidean distance in this case).

x_1	x_2	x_3	x_4	x_5	x_6
(-1, 1)	(-1, 1.5)	(1, 1)	(1, 2)	(2, 1.5)	(0, 4)

Table 3.1 Artificial data for a toy example of agglomerative hierarchical clustering.

1. Let $g_i = \{x_i\}$, $1 \leq i \leq 6$, be singleton classes.

2. The closest two grains/clusters are g_1 and g_2 (with the distance 0.5), so merge g_1 and g_2 as a new grain/cluster $g_{12} = \{x_1, x_2\}$ resulting in 5 grains/clusters g_{12}, g_3, g_4, g_5, g_6.

3. The closest two grains/clusters are g_3 and g_4 (with the distance 1) from the last step, so merge g_3 and g_4 into a new grain/cluster $g_{34} = \{x_3, x_4\}$ resulting in 4 clusters g_{12}, g_{34}, g_5, g_6.

4. The closest two grains/clusters are g_{34} and g_5 (with the distance 1.12) from the last step, so merge g_{34} and g_5 as a new grain/cluster $g_{345} = \{x_3, x_4, x_5\}$ resulting in 3 clusters g_{12}, g_{345}, g_6.

5. The closest two grains/clusters are g_{12} and g_{345} (with distance 3.04) from the last step, so merge g_{12} and g_{345} as a new grain/cluster $g_{12345} = \{x_i, 1 \leq i \leq 5\}$ resulting in 2 clusters g_{12345}, g_6.

6. Merge g_{12345} and g_6 into a new grain/cluster g_{123456} resulting in a single grain/cluster containing all data points.

This process obviously produces a clustering for each possible number of clusters of N data points, N through 1.

3.4 Application to Real Data with Discussions

3.4.1 Non-smoothed Grain Maps

We apply the agglomerative hierarchical method to orienlocations with distance defined in Chapter 3.2.5. The real data used here are orienlocations scanned at 3449 locations (with some missing data points) on the surface of a nickel specimen, which were collected in an experiment on EBSD machine precision (cf. Bingham et al., 2009a). The Euclidean distances between locations are scaled as suggested in Chapter 3.2.4. One can tune parameters w, c_o, a_o, c_l, a_l (defined in Chapter 3.2.5) and n (the number of clusters/grains) to produce sensible grain maps. The complete linkage (see Chapter 3.3) is used to produce grain maps in this manuscript. (Surely one could also change the linkage in the hierarchical clustering method to produce different grain maps.) Figure 3.1 is a grain map produced with $w = 1000$, $c_o = 4$, $a_o = 0$, $c_l = 4$, $a_l = 0$ and $n = 25$ (indicated by parameters in parentheses above the plot). The blank points in Figure 3.1 correspond to missing data points. As one can see, Figure 3.1

contains many singleton points surrounded by points belonging to other grains, which physically does not make much sense. These aberrant points are likely due to measurement errors in the process of EBSD scanning, and it is therefore desirable to"smooth out" such singleton points in the grain map.

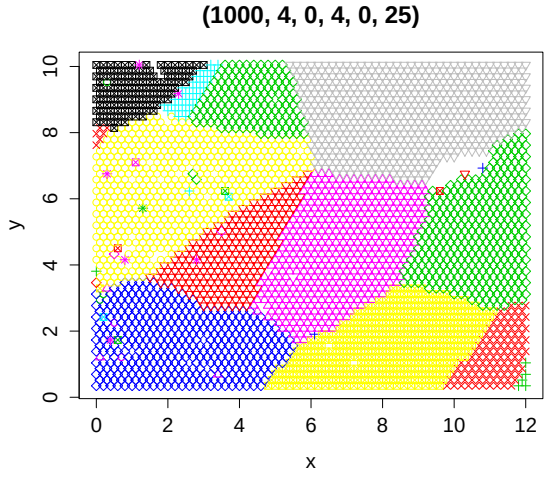

Figure 3.1 Grain map with $w = 1000$, $c_o = 4$, $a_o = 0$, $c_l = 4$, $a_l = 0$ and $n = 25$.

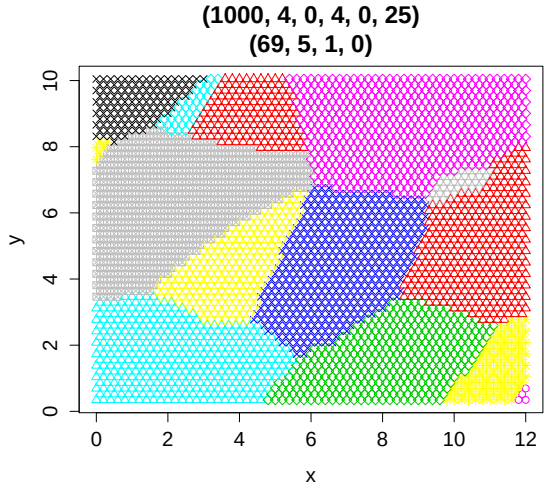

Figure 3.2 Smoothed grain map with $w = 1000$, $c_o = 4$, $a_o = 0$, $c_l = 4$, $a_l = 0$ and $n = 25$.

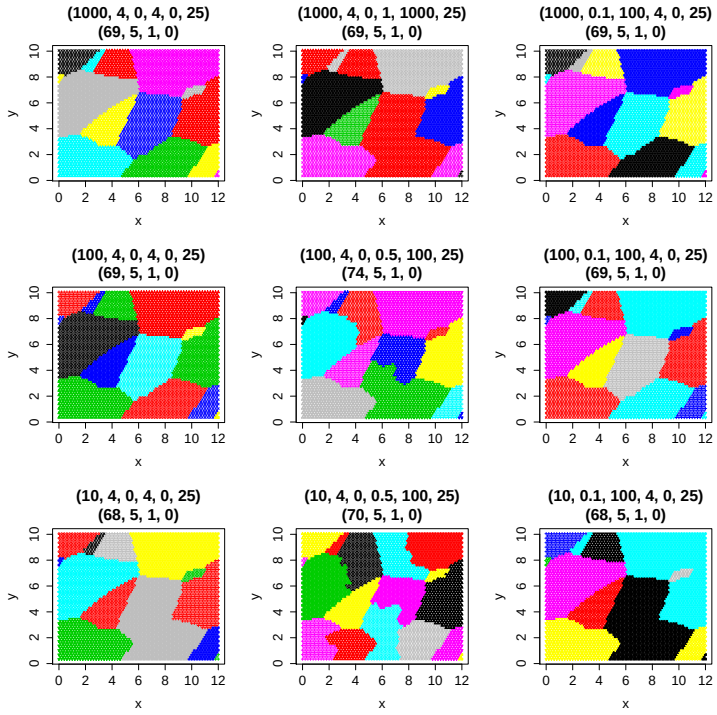

Figure 3.3 Each grain map corresponds to a different factor combination (ω, c_o, a_o, c_l, a_l, n) (the first line of parameters above plots), where $n = 25$ is the maximal number of clusters allowed. The 3 rows of grain maps correspond to ω of 1000, 100 and 10 respectively. The 3 columns of grain maps penalize no distances, the Euclidean distances and the orientation distances respectively. The second line of parameters above each plot contains the number of points smoothed out in each smoothing round.

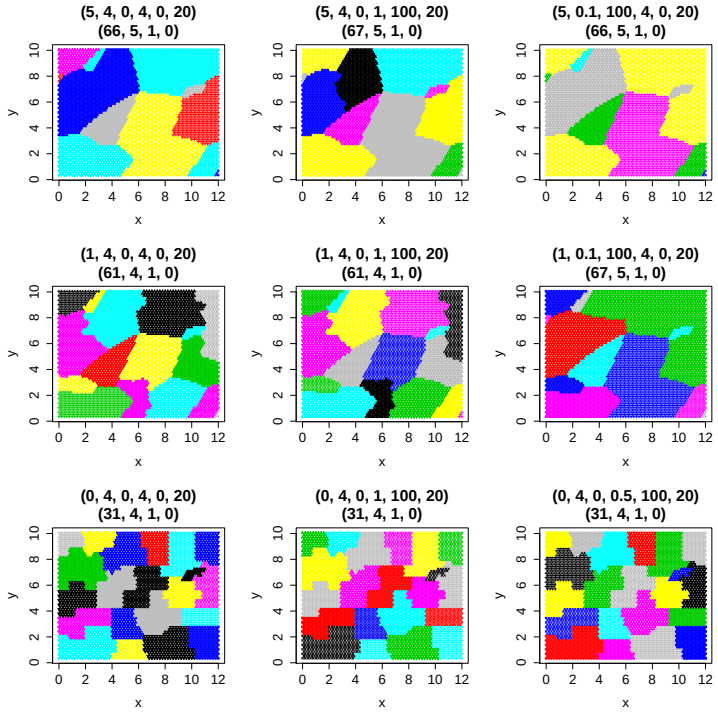

Figure 3.4 Each grain map corresponds to a different factor combination (ω, c_o, a_o, c_l, a_l, n) (the first line of parameters above plots), where $n = 20$ is the maximal number of clusters allowed. The 3 rows of grain maps correspond to ω of 5, 1 and 0 respectively. The 3 columns of grain maps penalize no distances, the Euclidean distances and the orientation distances respectively, except the bottom right one. The second line of parameters above each plot contains the number of points smoothed out in each smoothing round.

3.4.2 Smoothed Grain Maps

Let p be a 2-dimensional point. The ϵ-neighborhood of p (represented by $N_\epsilon(p)$) is the set of points (excluding p itself) within (Euclidean) distance ϵ from p, i.e.,

$$N_\epsilon(p) = \{p' : 0 < d_l(p, p') < \epsilon\}. \tag{3.15}$$

Let $n_\epsilon(p)$ be the number of elements in $N_\epsilon(p)$ and $k_\epsilon(p)$ be the number of elements in the same grain as p in $N_\epsilon(p)$. We can "smooth" the grain map using the following process.

(a) For a given grain map, put all missing data points into a new grain.

(b) For a point p, calculate $N_\epsilon(p)$ and $k_\epsilon(p)$ based on the grain map in step (a).

(c) If $k_\epsilon(p)/N_\epsilon(p) < \delta$ for a pre-chosen $\delta \in (0, 1)$, assign p to the grain with the greatest number of elements in $N_\epsilon(p)$.

(d) If there is a tie between two grains G_1 and G_2 in step (c) , i.e., the number of elements in $E_1 = G_1 \cap N_\epsilon(p)$ and $E_2 = G_2 \cap N_\epsilon(p)$ are the same, then break the tie using $d_g(p, G_1)$ and $d_g(p, G_2)$ (with the smaller one preferred), where d_g is as in Chapter 3.3 (i.e., choose the largest grain that is closest to p in Euclidean distance).

(e) Repeat (b)-(d) for each point in the grain map.

(f) Update the grain map.

(g) Repeat (b)-(f) r (say 10) times or until the grain map does not change.

Factors ϵ and δ in the above process control how much the grain map is smoothed. For greater ϵ and δ, the grain map is more smoothed out, and it is smoothed less for small ϵ and δ. In this manuscript, ϵ is chosen so that $N_\epsilon(p)$ has at most 7 points and δ is chosen to be 0.2. This allows singleton points (surrounded by points belonging to other grains) to be smoothed out; however, small clusters/grains with several (say more than 5) points will not be smoothed out.

Figure 3.2 is the smoothed version of Figure 3.1. The second line of parameters in parentheses above the plot means that the smoothing process took 4 rounds, and the number of

points smoothed out in the 4 rounds are 69, 5, 1 and 0 respectively. As one can see, singleton points (surrounded by points belonging to other grains) in Figure 3.1 are smoothed out, resulting a better-looking grain map. Figure 3.3 and Figure 3.4 present some smoothed grain maps with different choices of parameters. Figure 3.3 allows at most 25 clusters. The 3 rows of grain maps in Figure 3.3 correspond to ω (weight of squared orientation distance) of 1000, 100 and 10 respectively. The first column of grain maps in Figure 3.3 does not penalize (squared) orientation or Euclidean distances; the second column of grain maps in Figure 3.3 penalizes the Euclidean distance but not the orientation distance; the last column of grain maps in Figure 3.3 penalizes the orientation distance but not the Euclidean distance. Figure 3.4 allows at most 20 clusters. The 3 rows of grain maps in Figure 3.4 correspond to ω (weight of squared orientation distance) of 5, 1 and 0 respectively. The settings of penalization of distances in Figure 3.4 is similar to that in Figure 3.3 except the bottom right grain map. Since $\omega = 0$ for this case, the clustering-smoothing algorithm relies solely on the location information to make grain maps and it does not matter how the orientation distance is penalized. These plots give one some sense about how a real grain map changes according to different choices of parameters in our clustering-smoothing algorithm. These parameters "interact" with each other to produce different grain maps. Generally speaking, we observe the following patterns.

- Bigger ω (relying more on orientation information) tends to produce fewer and bigger grains on the grain map, and smaller ω produces more and smaller grains. This is because big ω often results in some very small grains which are smoothed out in the smoothing process.

- Penalizing the Euclidean distance (using smaller c_l and bigger a_l) tends to produce more and smaller grains on the grain map. This is intuitively plausible, as penalizing the Euclidean distance makes points far away from each other on the specimen surface unlikely to be in a same grain.

- Penalizing the orientation distance (using smaller c_o and bigger a_o) tends to produce fewer and bigger grains, which may initially sound counter-intuitive. The behavior occurs because penalizing the orientation distance makes clustering rely more on the orientation

information and thus produce fewer and bigger grains as discussed before.

3.5 A Simulation Study of the Clustering Algorithm

Generally speaking, it is hard to assess the goodness of clustering algorithms based on real data because we do not know what is the "right" grain map for the algorithm to produce. In this section, we assess plausibility of the clustering algorithm based on a simulation study. We generate orientations (rotation matrices) from the symmetric matrix Fisher von Mises distribution. The symmetric matrix Fisher von Mises distribution has 2 parameters, the "central" orientation S and the concentration parameter κ (with concentration increasing or variability decreasing in increasing κ). We denote the distribution as SMF(S, κ) in the following. The SMF(S, κ) distribution is a special case of the uniform-axis random-spin (UARS) model introduced in Bingham et al. (2009a) and is well studied in Du et al. (2013). A constructive way to generate orientations (rotation matrices) from the SMF(S, κ) distribution is introduced in Du et al. (2013).

To make it easier for the human eye to judge maps, we divide the locations where orientations were measured on the surface of an hypothetical specimen like the nickel specimen mentioned in Chapter 3.4.1 into 4×4 (approximate) rectangular areas with about equal sizes (i.e., the resulting different areas/hypothetical grains constitute a checkerboard pattern). At each location in rectangular area $i \in \{1, \ldots, 16\}$, an orientation (rotation matrix) from the SMF(S_i, κ_{in}) distribution is generated, where S_i is the central orientation for observations in the i^{th} area and k_{in} corresponds to the concentration parameter for these observations. The "central" orientations S_i, $i \in \{1, \ldots, 16\}$, are generated from the SMF(I_3, κ_{out}) distribution, where I_3 is the 3×3 identity matrix and k_{out} denotes the concentration parameter in generating the central orientations S_i in this hierarchical data formulation. For various level combinations of $\kappa_{out} \in \{10, 5, 3\}$ and $\kappa_{in} \in \{10, 20, 50\}$, the best (visually closest to the original 4×4 checkerboard) grain maps produced and the corresponding parameters producing the grain maps are shown in Figure 3.5. The 3 rows of grain maps in Figure 3.5 correspond to κ_{out} of 10, 5 and 3, and the 3 columns of grain maps correspond to κ_{in} of 10, 20 and 50. The top left grain map in Figure 3.5 does not capture most of the boundaries of the checkerboard. This is not surprising

as κ_{out} is no larger than κ_{in}, so that the resulting orientations contribute very little information to separate the areas/grains (i.e., in terms of orientations the grains are all distributionally equivalent) and the clustering algorithm has to rely mainly on the location information to cluster orienlocations. The grain maps in Figure 3.5 look better as one moves from left to right and from top to bottom. This makes sense because, for a fixed κ_{out}, it becomes easier to separate orienlocations when the orientations in each area/grain becomes more concentrated; and for a fixed κ_{in}, it is easier to separate orienlocations when the "central" orientations are spread out and thereby differentiated (smaller κ_{out} generates more vriable "central" orientations). The tuning parameters in the first line above each grain map correspond to ω, c_o, a_o, c_l, a_l and n respectively as explained in Chapter 3.4.1. We observe that ω (the weight of orientation distance) increases from left to right and from top to bottom in Figure 3.5. This means that the algorithm relies more and more on the orientation information when the orientations provide better information to separate different areas/grains, which makes sense in the application. The second line in the title in the grain maps contains the numbers of points that are smoothed out in each step. We can see that as the areas/grains are easier to separate, there are fewer points requiring additional smoothing. We also observe (not shown in Figure 3.5) that the ranges of optimal parameters become wider as the areas/grains becomes easier to separate.

3.6 Conclusion

This manuscript proposes a hierarchical clustering method for clustering (equivalence classes of) orientations with spatial location information. A reasonable and flexible metric was defined for orienlocations (data points in the form of an orientation paired with a spatial location). As a remedy for handling missing or outlying data points, a smoothing method was also proposed to refine and produce better-looking grain maps. The clustering-smoothing algorithm has several tuning parameter "knobs" that can be tuned or adjusted to produce desired grain maps. The simulation study in Chapter 3.5 demonstrated that the clustering algorithm behaves in a principled manner with respect to the informational content in the data and can correctly identify clusters of orienlocations. The clustering-smoothing algorithm introduced in this manuscript is more logically coherent than ad hoc clustering methods in current real use,

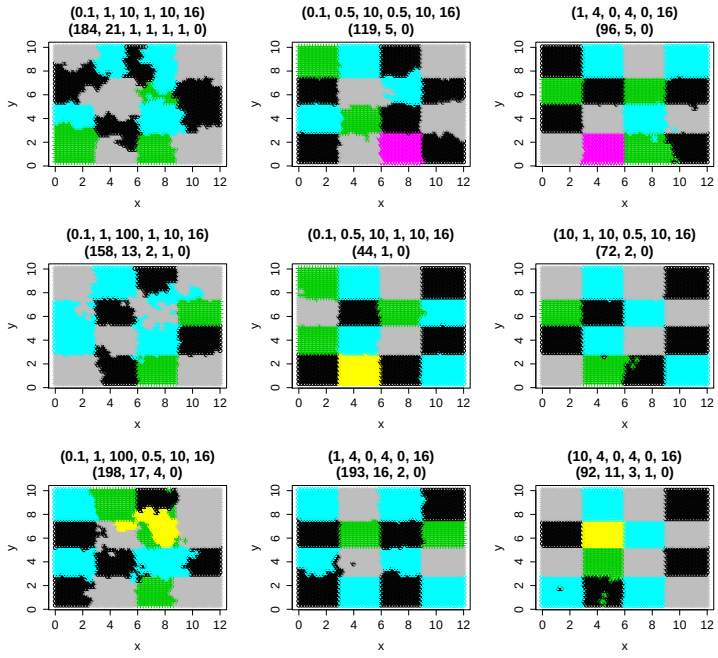

Figure 3.5 Best grain maps and corresponding parameters used to produce these grain maps for different combinations of κ_{out} and κ_{in}. The 3 rows of grain maps correspond to κ_{out} of 10, 5 and 3, and the 3 columns of grain maps correspond to κ_{in} of 10, 20 and 50.

and is potentially helpful in texture analysis. Implementation of the methods here allowing a user to adjust tuning parameters and thus the appearance of a map for his or her specimen offers great promise for the advance of the science.

CHAPTER 4. A CLASS OF STATIONARY AND ERGODIC MARKOV CHAINS DEFINED ON PARTITIONS OF A FINITE SET WITH APPLICATIONS IN BAYESIAN CLUSTERING

A paper submitted to the Journal of Bayesian Analysis

Chuanlong Du[1]

Abstract

An interesting class of Markov chains with state spaces partitions of a finite set is studied in this manuscript. These Markov chains have some properties that make them attractive in their own right. An important area of potential application for these processes is Bayesian model-based clustering, which is illustrated in Chapter 4.4.

4.1 Introduction

In this manuscript, an interesting class of Markov chains (that we call the Du processes) with state spaces that are partitions of a finite index set is studied. In Chapter 4.2, a formal definition of the Du processes is given. It is shown in Chapter 4.3 that the Du processes have some attractive properties.

An application of the Du processes in Bayesian model-based clustering is illustrated in Chapter 4.4. The new approach they provide to Bayes clustering is easier to understand and implement than other (latent-class-variables-based) methods, such as the Bayesian finite mixture model approach and the Bayesian Dirichlet mixture model approach. It also provides a control over the jumping rate of the partition parameter (which affects the per iteration

[1]Graduate student, Department of Statistics, Iowa State University

quality of posterior draws) in a Metropolis algorithm. In latent-class-variables-based methods, though many Metropolis-Hasting algorithms have been proposed, few of them offer a clear way to control jumping rates of the latent class variables.

4.2 Definition of Markov Chains on Partitions

Let $\Omega = \{1, 2, \ldots, n\}$ be a set of n elements. A collection of sets $\mathcal{P} = \{S_1^{\mathcal{P}}, S_2^{\mathcal{P}}, \ldots, S_{k(\mathcal{P})}^{\mathcal{P}}\}$, of size $k(\mathcal{P})$ is called a partition of Ω if $S_i^{\mathcal{P}} \cap S_j^{\mathcal{P}} = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{k} S_i^{\mathcal{P}} = \Omega$. Let $\mathcal{A}(\Omega)$ be the collection of all partitions of Ω.

Definition 1. *Let G be a distribution on Ω with probability mass $g_i > 0$ on i, for all $i \in \Omega$. A 1-step random mapping T_G^1 on $\mathcal{A}(\Omega)$ is one such that $\forall \mathcal{P} = \{S_1^{\mathcal{P}}, S_2^{\mathcal{P}}, \ldots, S_{k(\mathcal{P})}^{\mathcal{P}}\} \in \mathcal{A}(\Omega)$, $T_G^1(\mathcal{P})$ is random element of $\mathcal{A}(\Omega)$ determined as follows:*

1. *Generate i from G, and find the element of $\mathcal{P}$ to which i belongs, call it $S_{j(i)}^{\mathcal{P}}$.*
2. *If $S_{j(i)}^{\mathcal{P}}$ is a singleton, i.e., $S_{j(i)}^{\mathcal{P}} = \{i\}$, move i to one of the sets $S_l^{\mathcal{P}}$ for $l \neq j(i)$ chosen at random with equal probabilities $1/(k(\mathcal{P}) - 1)$.*
3. *If $S_{j(i)}^{\mathcal{P}}$ is not a singleton, i.e., $S_{j(i)}^{\mathcal{P}}$ contains element(s) other than i, move i to one of the sets $S_l^{\mathcal{P}}$ for $l \neq j(i)$ or to a new singleton set $\{i\}$, where the choice is made at random with equal probabilities $1/k(\mathcal{P})$.*

For convenience, we also denote T_G^1 as T_G in the balance of this manuscript.

Definition 2. *Let G be as in Definition 1. A k-step ($k > 1$) random mapping T_G^k on $\mathcal{A}(\Omega)$ is the composition of k maps T_G, i.e., $T_G^k(\mathcal{P}) = T_G(T_G^{k-1}(\mathcal{P}))$,*
$\forall \mathcal{P} \in \mathcal{A}(\Omega)$.

Definition 3. *Let G be as in Definition 1, and F be (possibly degenerate) cdf putting mass on positive integers. An F distributed random mapping T_G^F on $\mathcal{A}(\Omega)$ is such that $Pr(T_G^F = T_G^x) = F(x) - F(x-1), \forall x \in \mathbb{N}_+$, where $\mathbb{N}_+$ is the set of positive integers.*

Note that when F is degenerate at $k(\geq 1)$, T_G^F is the same as T_G^k. We will use $T_G^{F,m}$ to stand for the composition of m maps T_G^F in the balance of this manuscript. Note that $T_G^{k,m} = (T_G^k)^m = \left((T_G)^k\right)^m = (T_G)^{km} = T_G^{km}$.

Definition 4. *Let G and F be as in Definition 3. A Markov chain $\{X_m\}_{m=0}^{\infty}$ with state space $\mathcal{A}(\Omega)$ is called a Du(n,G,F) process, if for all $m \in \mathbb{N}$, the conditional distribution of X_{m+1} given X_m is that of $T_G^F(X_m)$, i.e., $X_{m+1}|X_m \stackrel{d}{=} T_G^F(X_m)$.*

From the definition, the Du(n,G,F) process is a stationary Markov chain. It turns out that when Ω has at least 3 elements (i.e., $n \geq 3$) and G and F are as in Definition 3, the Du(n,G,F) process is ergodic with stationary distribution the uniform distribution on $\mathcal{A}(\Omega)$. When we use notation Du(n,G,F) in the balance of this manuscript, we always assume that $n \geq 3$ and that G and F are as in Definition 3. When the distribution of F is degenerate at k, we also denote the process as Du(n,G,k).

In the balance of this section, we prepare some terminology that will be useful in establishing properties of the Du processes.

Definition 5. *Let G and F be as in Definition 3. A sequence of partitions $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$ ($k \geq 1$) is called a k-step path of T_G^F from $\mathcal{P}_0$ to $\mathcal{P}_k$, if $\prod_{i=1}^{k} Pr(T_G(\mathcal{P}_{i-1}) = \mathcal{P}_i) > 0$ and $F(k) - F(k-1) > 0$. Denote this as $T_G^F : \mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$.*

Definition 6. *Let G and F be as in Definition 3. A k-step ($k \geq 1$) path of T_G^F from $\mathcal{P}_0$ to $\mathcal{P}_k$, $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$ is reversible under mapping T_G^F if $T_G^F : \mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0$, or equivalently $\prod_{i=1}^{k} Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) > 0$.*

We'll show later (see Proposition 4.3.7) that $\forall k \in \mathbb{N}_+$, $\mathcal{P}, \mathcal{P}' \in \mathcal{A}(\Omega)$ and any random mapping T_G^F, the k-step paths (if any) of T_G^F from $\mathcal{P}$ to $\mathcal{P}'$ are reversible under T_G^F. For this reason, we call path $\mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0$ the reverse of path $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$.

Definition 7. *Let G and F be as in Definition 3. A k-step ($k \geq 1$) path of T_G^F from $\mathcal{P}_0$ to $\mathcal{P}_k$, $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$ is symmetric under mapping T_G^F if it's reversible under mapping T_G^F and $\prod_{i=1}^{k} Pr(T_G(P_{i-1}) = P_i) = \prod_{i=1}^{k} Pr(T_G(P_i) = P_{i-1})$.*

4.3 Some Important Properties of the Du(n,G,F) Process

Theorem 4.3.1. *The Du(n,G,F) process is ergodic.*

Theorem 4.3.2. *The transition function of the Du(n,G,F) process is doubly stationary.*

Theorem 4.3.3. *The Du(n,G,F) process has a unique stationary distribution which is the uniform distribution on* $\mathcal{A}(\Omega)$.

Here are skeletons of proofs of the three theorems. The detailed proofs of the propositions in this section can be found in the appendix.

Theorem 4.3.1 follows immediately from the following 3 propositions.

Proposition 4.3.4. *The Du(n,G,F) process is irreducible.*

Proposition 4.3.5. *The Du(n,G,F) process is positive recurrent.*

Proposition 4.3.6. *The Du(n,G,F) process is aperiodic.*

To prove Theorem 4.3.2, we need the following two propositions.

Proposition 4.3.7. *Any k-step* ($k \geq 1$) *path of* T_G^F *is reversible and symmetric under* T_G^F.

Proposition 4.3.8. *Let* $M_{G,F}^k(\mathcal{P}, \mathcal{P}') = \{T_G^F : \mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k | \mathcal{P}_0 = \mathcal{P} \text{ and } \mathcal{P}_k = \mathcal{P}'\}$ *be the collection of all k-step paths of random mapping* T_G^F *from* $\mathcal{P}$ *to* $\mathcal{P}'$, *and* $W_{G,F}^k(\mathcal{P}, \mathcal{P}') = \{\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k | \mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0 \in M_{G,F}^k(\mathcal{P}, \mathcal{P}')\}$, *i.e., the collection of reverse of paths in* $M_{G,F}^k(\mathcal{P}, \mathcal{P}')$. *Then* $W_{G,F}^k(\mathcal{P}, \mathcal{P}') = M_{G,F}^k(\mathcal{P}', \mathcal{P})$, *i.e., the collection of reverse k-step paths of* T_G^F *from* $\mathcal{P}$ *to* $\mathcal{P}'$ *is the same as the collection of k-step paths of* T_G^F *from* $\mathcal{P}'$ *to* $\mathcal{P}$.

Let $f(k) \equiv F(k) - F(k-1)$, and

$$Pr(M_{G,F}^k(\mathcal{P}, \mathcal{P}')) \equiv f(k) \sum_{\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k \in M_{G,F}^k(\mathcal{P}, \mathcal{P}')} \prod_{i=1}^{k} Pr(T_G(\mathcal{P}_{i-1}) = \mathcal{P}_i). \tag{4.1}$$

From Proposition 4.3.7, the last term in (4.1) is

$$
\begin{aligned}
& f(k) \sum_{\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k \in M^k_{G,F}(\mathcal{P},\mathcal{P}')} \prod_{i=1}^{k} Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) \\
& = f(k) \sum_{\mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0 \in W^k_{G,F}(\mathcal{P},\mathcal{P}')} \prod_{i=1}^{k} Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) \\
& = f(k) \sum_{\mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0 \in M^k_{G,F}(\mathcal{P}',\mathcal{P})} \prod_{i=1}^{k} Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) \\
& = Pr(M^k_{G,F}(\mathcal{P}',\mathcal{P})),
\end{aligned}
$$

i.e., $Pr(M^k_{G,F}(\mathcal{P},\mathcal{P}')) = Pr(M^k_{G,F}(\mathcal{P}',\mathcal{P}))$. So,

$$
\begin{aligned}
Pr(T^F_G(\mathcal{P}) = \mathcal{P}') &= \sum_{k \ni M^k_{G,F}(\mathcal{P},\mathcal{P}') \neq \emptyset} Pr(M^k_{G,F}(\mathcal{P},\mathcal{P}')) \\
&= \sum_{k \ni M^k_{G,F}(\mathcal{P}',\mathcal{P}) \neq \emptyset} Pr(M^k_{G,F}(\mathcal{P}',\mathcal{P})) \\
&= Pr(T^F_G(\mathcal{P}') = \mathcal{P})
\end{aligned}
$$

i.e., the transition function of the Du(n,G,F) process is doubly stationary.

Theorem 4.3.3 follows from Theorems 4.3.1 and 4.3.2. The Du(n,G,F) process is ergodic, so it has a unique stationary distribution. Since the transition mechanism of the Du(n,G,F) process is doubly stationary,

$$
\sum_{\mathcal{P} \in \mathcal{A}(\Omega)} Pr(T^F_G(\mathcal{P}) = \mathcal{P}') = \sum_{\mathcal{P} \in \mathcal{A}(\Omega)} Pr(T^F_G(\mathcal{P}') = \mathcal{P}) = 1,
$$

i.e., the elements of each column of the transition matrix (denote this matrix as E) of the Du(n,G,F) process sum to 1. Let K be the number of partitions in $\mathcal{A}(\Omega)$ (the so-called "Bell number") and $\boldsymbol{c} = (1/K, \ldots, 1/K)$, $\boldsymbol{c}E = \boldsymbol{c}$. That is, the uniform distribution on $\mathcal{A}(\Omega)$ is the stationary distribution of the Du(n,G,F) process.

4.4 An Illustration of the Use of the Du(n,G,F) Process in Bayesian Clustering

There are basically two approaches to Bayesian model-based clustering in literature, Bayesian finite mixture modeling (see Nobile and Fearnside, 2007) and Bayesian Dirichlet mixture modeling (see Lau and Green, 2007; Dubey et al., 2004, etc). The Bayesian Dirichlet mixture model is equivalent to a Bayesian infinite mixture model, and can thus be thought of as a generalization of the Bayesian finite mixture model. The Dirichlet process is closely related to many other processes. The Chinese restaurant process and the Polya urn scheme are conditional views of the Dirichlet process; the stick-breaking process is a convenient way for constructing the Dirichlet process; the Pitman-Yor process (Pitman and Yor, 1997) is a generalization of the Dirichlet process which has more flexible tail behavior. Because of its flexibility and the possibility of using conjugate priors, the Dirichlet process is widely used in Bayesian clustering methods.

Both the Bayesian finite mixture model and the Bayesian Dirichlet process approaches to clustering introduce latent class variables for the observations. Markov chains on the latent class variables (and other parameters if any) are then simulated using either Gibbs or Metropolis-Hasting algorithms (see Neal, 2000; Nobile and Fearnside, 2007; Jain and Neal, 2004, etc). This latent-class-variables representation of clusters is not intrinsic. Instead of Markov chains on latent variables, a more natural, convenient, and potentially better approach is to consider partitions of the n observations directly. It's easy to understand and invites use of a simpler Metropolis algorithm for making posterior draws (e.g. based on proposals from the Du(n,G,F) process transition mechanism). While other approaches have been well-studied, the direct approach is not common in the statistics literature.

To illustrate, let $y_1, \ldots, y_n$ be independent observations, and $\mathcal{P} = \{S_1, \ldots,$
$S_k\}$ be a partition of the index set. We assume that conditioned on $\mathcal{P}$, observations are independent,

$$y_i \sim h(.|\theta_S), \forall i \in S,$$

where S is a set in partition $\mathcal{P}$. At present, let's assume that the parameters θ_S can be

integrated out so that the partition is the only parameter left in the model. Since the total number of partitions (the Bell number) is extremely large for n of a size typical in clustering problems, it's often practically impossible to sum over all possible partitions for purposes of evaluating posterior probabilities. A potentially computationally tractable way to explore the posterior of $\mathcal{P}$ is to make draws from it using Metropolis-Hasting algorithms. The fact that the transition function of the Du(n,G,F) is doubly stationary makes $T_G^F(\mathcal{P})$ a good choice for a proposal. Since the proposal distribution is symmetric, a simple Metropolis algorithm results. Generally speaking, the more indexes we move when proposing a partition, the "further away" a proposed partition from a current partition, and thus on average the smaller is the jumping rate. This suggests a way to control the jumping rate. The distribution F serves as a tuning parameter for a Metropolis algorithm that should control the jumping rate.

After a Markov chain on partitions burns in, some indexes are more unstable/active (produce large acceptance ratios when moved between clusters) than others. It's probably a good idea for these unstable/active points to have a relatively higher probabilities to be chosen to move. This will increase the jumping rate on average. So the distribution G serves as another tuning parameter for a Metropolis algorithm based on the Du(n,G,F) process transition mechanism that can be varied looking for a good jumping rate.

As a simple illustration, I independently simulated 5 observations from N(4,2), 8 observations from N(9,2), and 7 observations from N(14,2), where N(μ,σ^2) is the normal distribution with mean μ and variance σ^2. The 20 observations are presented in Table 4.1. Let y_i be

-0.43	3.31	8.54	3.01	2.33	11.70	6.79	8.65	9.90	8.17
10.72	10.94	11.06	11.67	14.70	10.03	16.35	9.82	18.77	12.61

Table 4.1 Simulated data from normal distributions.

centered (by overall mean) versions of the observations. The model I use for clustering purposes will be that conditioned on $\mathcal{P}$ observations y_i are independent

$$y_i \sim \mathrm{N}(\mu_S, \sigma^2), \forall i \in S,$$

where S is a class in $\mathcal{P}$. For convenience, σ was treated as a known constant. However, the

particular value used for σ is not critical. As the prior for each μ_S, $\mathrm{N}(0, \epsilon^2)$ was used. As the prior over partitions,

$$\pi(\mathcal{P}) \propto \exp\{-FP(\mathcal{P}; \alpha, \beta)\}$$

was used, where $FP(\mathcal{P})$ is what might be termed a "fragment penalty" which puts penalty on small classes (especially singletons), and has the form

$$FP(\mathcal{P}; \alpha, \beta) = \sum_{S \in \mathcal{P}} \alpha \left(\frac{1}{k_S}\right)^{\beta}.$$

Integrating μ_S's out, it can be shown that the posterior distribution over partitions is

$$p(\mathcal{P}) \propto \prod_{S \in \mathcal{P}} \sqrt{\frac{R}{R + k_S}} \exp\left\{\frac{\left(\sum_{i \in S} y_i\right)^2}{2\sigma^2(k_S + R)}\right\} \exp\{-FP(\mathcal{P}; \alpha, \beta)\},$$

where $R = \frac{\sigma^2}{\epsilon^2}$. To sample from this posterior distribution, the following Metropolis algorithm can be used:

1. In step i, propose $\mathcal{P}^* = T_G^F(\mathcal{P}^{(i)})$.

2. Calculate acceptance ratio $r = \frac{p(\mathcal{P}^*)}{p(\mathcal{P}^{(i)})}$, and accept $\mathcal{P}^*$ with probability $min\{r, 1\}$.

In this illustrative example, the parameters used were $\sigma = 2$, $R = 0.01$, $\alpha = 6$, $\beta = 1$, G is $U(\{1, \ldots, n\})$ and F is the distribution of $X + 1$ for $X \sim Poisson(0.01)$. 100,000 draws were made after a burn-in of 10,000 iterations starting from the partition with a single class. The partition with highest estimated posterior probability was

$$\{\{0, 1, 3, 4\}, \{2, 6, 7, 9, 17\}, \{5, 8, 10, 11, 12, 13, 15, 19\}, \{14, 16, 18\}\}$$

Starting from the means of observations with indexes in the above groups, a k-means algorithm (with $k = 4$) yields the same result.

4.5 Conclusion

The Du(n,G,F) process is an appealing probabilistic structure in its own right, but its most obvious potential for contributing to statistical practice is in the realm of Bayesian model-based clustering. Bayesian model-based clustering using a direct partition representation is more

natural and intuitively appealing than methods built on latent class variables. The transition scheme of the Du(n,G,F) process can be an element of a simple Metropolis algorithm for Bayesian model-based clustering using a direct partition representation. It provides convenient ways for tuning jumping rates to improve the mixing of the Markov chain. It's possible to develop Metropolis algorithms for Bayesian clustering problems more complex than the simple illustration provided in this manuscript.

CHAPTER 5. GENERAL CONCLUSIONS

Chapter 2 demonstrates that models on equivalence classes of rotations provide a logically coherent way to treat data arising in some materials science applications. This approach does not require any *ad hoc* preprocessing of "unlabeled data" into orientations, which does not work in some situations. An effective Bayesian method for one-sample inference is developed for the new UARS($[\boldsymbol{S}], \kappa$) models on rotation matrix equivalence classes introduced in Chapter 2. The Bayesian method (for the SMF($[\boldsymbol{S}], \kappa$) model) is easier to interpret and generally works better than maximum likelihood-based methods especially when the sample size is small.

Chapter 3 proposes a hierarchical clustering method for clustering equivalence classes of orientations with spatial location information. A reasonable and flexible metric was defined for orienlocations (data points in the form of an orientation paired with a spatial location). As a remedy for handling missing or outlying data points, a smoothing method was also proposed to refine and produce better-looking grain maps. The clustering-smoothing algorithm has several tuning parameter "knobs" that can be tuned or adjusted to produce desired grain maps. It is demonstrated via simulation in Chapter 3.5 that the clustering algorithm behaves in a principled manner with respect to the informational content in the data and can correctly identify clusters of orienlocations. The clustering-smoothing algorithm introduced in Chapter 3 is more logically coherent than ad hoc clustering methods in current real use, and is potentially helpful in texture analysis. The algorithm allows users to adjust tuning parameters and thus the appearance of a map for their specimen offers great promise for the advance of the science.

Chapter 4 studies the Du(n,G,F) process, which is an appealing probabilistic structure in its own right. The most obvious potential for contributing to statistical practice of the Du(n,G,F) process is in the realm of Bayesian model-based clustering. Bayesian model-based clustering using a direct partition representation is more natural and intuitively appealing than methods

built on latent class variables. The transition scheme of the Du(n,G,F) process can be an element of a simple Metropolis algorithm for Bayesian model-based clustering using a direct partition representation. It provides convenient ways for tuning jumping rates to improve the mixing of the Markov chain. Chapter 4 explores use of the Du(n,G,F) process only in simple situations. More work need to be done to apply it to general situations.

APPENDIX A. UARS($\boldsymbol{S}, \kappa$) MODELS FOR ROTATION MATRICES

Let Ω be the collection of all 3×3 rotation matrices, i.e., $SO(3)$. Bingham et al. (2009a) and Hielscher et al. (2010) identified a class of UARS models on Ω, which are useful for describing random rotations symmetrically distributed around a fixed, mean rotation (i.e., central location) parameter $\boldsymbol{S} \in \Omega$, and where the amount of variability in rotations can be directly controlled by a concentration parameter $\kappa > 0$ in the model. Rotations in the UARS model class have a single simple, geometric construction as follows. Suppose we have a unit vector $\boldsymbol{u} \in \mathbb{R}^3$, and we spin the axes of the standard coordinate system (represented by the columns of the identity matrix $\boldsymbol{I}_{3\times 3}$) around the direction $\boldsymbol{u}$ (i.e., a signed axis) counter-clockwise through angle r. Upon this rotation, the columns of $\boldsymbol{I}_{3\times 3}$ move to positions given by a corresponding rotation matrix

$$\boldsymbol{M}(\boldsymbol{u}, r) = \boldsymbol{u}\boldsymbol{u}^T + (\boldsymbol{I}_{3\times 3} - \boldsymbol{u}\boldsymbol{u}^T)\cos r + \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \sin r, \tag{A.1}$$

which is Euler's angle-axis representation of a 3-D rotation; see Figure 1.1 for illustration.

UARS models are a stochastic version of this angle-axis representation. If we now take $\boldsymbol{u}$ as uniformly distributed on the unit sphere and independent of a random angle $r \sim \text{Cir}(\kappa)$, where $\text{Cir}(\kappa)$ is a circular distribution on $(-\pi, \pi]$, symmetric around 0 and with a positive concentration parameter κ (with concentration increasing in κ), then we obtain a random rotation matrix denoted by $\boldsymbol{M}(\boldsymbol{u}, r)$ whose distribution we denote as $\text{UARS}(I, \kappa)$. For $\boldsymbol{S} \in \Omega$, $\boldsymbol{S} \cdot \boldsymbol{M}(\boldsymbol{u}, r) \sim \text{UARS}(\boldsymbol{S}, \kappa)$ gives a so-called UARS random rotation with location and concentration parameters respectively $\boldsymbol{S}$ and κ. Different choices of the circular distribution for r produce different models for symmetric random rotations. If the $\text{Cir}(\kappa)$ distribution for the angle r has a density $C(r|\kappa)$ (with respect to the usual Lebesgue measure), then a $\text{UARS}(\boldsymbol{S}, \kappa)$

rotation has a corresponding density

$$f(\boldsymbol{O}|\boldsymbol{S},\kappa) = \frac{4\pi}{3-\text{tr}(\boldsymbol{S}^T\boldsymbol{O})} C\left(\arccos(\frac{\text{tr}(\boldsymbol{S}^T\boldsymbol{O})-1}{2})|\kappa\right),\ \boldsymbol{O},\boldsymbol{S}\in\Omega,\ \kappa>0, \tag{A.2}$$

with respect to the "uniform distribution" on Ω (or Haar measure), which provides a dominating measure for defining densities on Ω (see Downs, 1972).

APPENDIX B. DETAILED PROOFS OF THEOREMS IN CHAPTER 4.3

Lemma B.0.1. *Let G be as in Definition 1. Any 1-step path of T_G is reversible and symmetric under T_G.*

Proof. Let $\mathcal{P} \in \mathcal{A}(\Omega)$ be a partition with k subsets. Without loss of generosity, suppose by applying T_G to $\mathcal{P}$, the element "1" is moved, and the resulting partition is $\mathcal{P}'$.

(i) If "1" is a singleton subset in $\mathcal{P}$, and "1" is moved to some existing subset $C \in \mathcal{P}$ with at least 2 elements (excluding the singleton $\{1\}$, there are $k-1$ subsets that "1" can move to), then partition $\mathcal{P}'$ has $k-1$ subsets including $C' = C \cup \{1\}$ which has at least 3 elements.

$$\begin{aligned} Pr(T_G(\mathcal{P}) = \mathcal{P}') &= Pr(\text{select "1"})Pr(\text{move "1" to } C) \\ &= g_1 \times \frac{1}{k-1} = \frac{g_1}{k-1} > 0. \end{aligned}$$

Applying T_G to partition $\mathcal{P}'$, if element "1" is selected, since it's in subset $C' \in \mathcal{P}'$ which has at least 3 elements, it's possible to move it to a new singleton subset (excluding C' and including a new singleton class, there are $k-1$ subsets that "1" can move to).

$$\begin{aligned} Pr(T_G(\mathcal{P}') = \mathcal{P}) &= Pr(\text{select "1"})Pr(\text{move "1" to a new singleton}) \\ &= g_1 \times \frac{1}{k-1} = \frac{g_1}{k-1} = Pr(T_G(\mathcal{P}) = \mathcal{P}') > 0. \end{aligned}$$

(ii) If "1" is a singleton subset in $\mathcal{P}$, and "1" is moved to some existing singleton subset $C \in \mathcal{P}$ (without loss of generality assume $C = \{2\}$, and as in (i) there are $k-1$ subsets that "1" can move to), then partition $\mathcal{P}'$ has $k-1$ subsets including $C' = \{1, 2\}$. Notice that by

moving "2" in $\mathcal{P}$ to the singleton subset $\{1\}$, we also obtain $\mathcal{P}'$. So

$$\begin{aligned} Pr(T_G(\mathcal{P}) = \mathcal{P}') &= Pr(\text{select "1"})Pr(\text{move "1" to } \{2\}) \\ &\quad + Pr(\text{select "2"})Pr(\text{move "2" to } \{1\}) \\ &= g_1 \times \frac{1}{k-1} + g_2 \times \frac{1}{k-1} = \frac{g_1 + g_2}{k-1} > 0. \end{aligned}$$

Applying T_G to partition $\mathcal{P}'$, if either "1" or "2" is selected, since $C' \in \mathcal{P}'$ is not a singleton, we can move the selected element to a new singleton subset (excluding C' and including a new singleton class, there are $k-1$ subsets that "1" or "2" can move to). So

$$\begin{aligned} Pr(T_G(\mathcal{P}') = \mathcal{P}) &= Pr(\text{select "1"})Pr(\text{move "1" to a new singleton}) \\ &\quad + Pr(\text{select "2"})Pr(\text{move "2" to a new singleton}) \\ &= g_1 \times \frac{1}{k-1} + g_2 \times \frac{1}{k-1} = \frac{g_1 + g_2}{k-1} \\ &= Pr(T_G(\mathcal{P}) = \mathcal{P}') > 0. \end{aligned}$$

(iii) If "1" is in a subset B of $\mathcal{P}$ with 2 elements (without loss of generosity, assume $B = \{1, 2\}$), and "1" is moved to a new singleton subset, by reversing the proof in (ii) we have

$$Pr(T_G(\mathcal{P}') = \mathcal{P}) = Pr(T_G(\mathcal{P}) = \mathcal{P}') = \frac{g_1 + g2}{k} > 0.$$

(iv) If "1" is in a subset $B \in \mathcal{P}$ with 2 elements (without loss of generality assume $B = \{1, 2\}$, and "1" is moved to some existing subset $C \in \mathcal{P}$ (excluding B and including a new singleton subset, there are k subsets that "1" can move to), the partition $\mathcal{P}'$ also has k subsets including $B' = \{2\}$ and $C' = C \cup \{1\}$ which has at least 2 elements.

$$\begin{aligned} Pr(T_G(\mathcal{P}) = \mathcal{P}') &= Pr(\text{select "1"})Pr(\text{move "1" to } C) \\ &= g_1 \times \frac{1}{k} = \frac{g_1}{k} > 0. \end{aligned}$$

Applying T_G to partition $\mathcal{P}'$, if "1" is selected, since $C' \in \mathcal{P}'$ is not a singleton, it's possible to move "1" to any subset other than C' in $\mathcal{P}'$ and a new singleton subset (excluding C' and including a new singleton, there are k subsets that "1" can move to).

$$\begin{aligned} Pr(T_G(\mathcal{P}') = \mathcal{P}) &= Pr(\text{select "1"})Pr(\text{move "1" to } B') = g_1 \times \frac{1}{k} \\ &= \frac{g_1}{k} = Pr(T_G(\mathcal{P}) = \mathcal{P}') > 0. \end{aligned}$$

(v) If "1" is in a subset $B \in \mathcal{P}$ with at least 3 elements, and "1" is moved to a new singleton subset, by reversing the proof in (i) we have

$$Pr(T_G(\mathcal{P}') = \mathcal{P}) = Pr(T_G(\mathcal{P}) = \mathcal{P}') = \frac{g_1}{k} > 0.$$

(vi) If "1" is in a subset $B \in \mathcal{P}$ with at least 3 elements, and "1" is moved to some existing subset $C \in \mathcal{P}$, as in (iv) we can show

$$Pr(T_G(\mathcal{P}') = \mathcal{P}) = Pr(T_G(\mathcal{P}) = \mathcal{P}') = \frac{g_1}{k} > 0.$$

To sum up, $\mathcal{P} \to \mathcal{P}'$ is reversible and symmetric under T_G. □

Proof of Proposition 4.3.7

Proof. Let $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$ be a k-step path of T_G^F from $\mathcal{P}_0$ to $\mathcal{P}_k$, from Definition 5,

$$\prod_{i=1}^{k} Pr(T_G(\mathcal{P}_{i-1}) = \mathcal{P}_i) > 0.$$

We have shown in Lemma B.0.1 that any 1-step path of T_G is reversible and symmetric under T_G, so,

$$Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) = Pr(T_G(\mathcal{P}_{i-1}) = \mathcal{P}_i) > 0, i = 1, \ldots, k.$$

Multiplying both left and right sides of this across values of i, we have

$$\prod_{i=1}^{k} Pr(T_G(\mathcal{P}_i) = \mathcal{P}_{i-1}) = \prod_{i=1}^{k} Pr(T_G(\mathcal{P}_{i-1}) = \mathcal{P}_i) > 0,$$

i.e., $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k$ is reversible and symmetric under T_G^F. □

Proof of Proposition 4.3.8

Proof. From Proposition 4.3.7, every path in $M_{G,F}^k(\mathcal{P}, \mathcal{P}')$ is reversible, so $\forall \mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k \in M_{G,F}^k(\mathcal{P}, \mathcal{P}')$, $\mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0 \in M_{G,F}^k(\mathcal{P}', \mathcal{P})$. This implies that $W_{G,F}^k(\mathcal{P}, \mathcal{P}') \subset M_{G,F}^k(\mathcal{P}', \mathcal{P})$.

Similarly, $\forall \mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k \in M_{G,F}^k(\mathcal{P}', \mathcal{P})$, $\mathcal{P}_k \to \mathcal{P}_{k-1} \to \ldots \to \mathcal{P}_0 \in M_{G,F}^k(\mathcal{P}, \mathcal{P}')$, which furthers implies $\mathcal{P}_0 \to \mathcal{P}_1 \to \ldots \to \mathcal{P}_k \in W_{G,F}^k(\mathcal{P}, \mathcal{P}')$. So $M_{G,F}^k(\mathcal{P}', \mathcal{P}) \subset W_{G,F}^k(\mathcal{P}, \mathcal{P}')$, and thus $W_{G,F}^k(\mathcal{P}, \mathcal{P}') = M_{G,F}^k(\mathcal{P}', \mathcal{P})$. □

The proof of Theorem 4.3.2 relies only on Propositions 4.3.7 and 4.3.8, so it's proven. Theorem 4.3.2 is helpful for proving the following lemma.

Lemma B.0.2. *The Du(n,G,1) process is irreducible.*

Proof. Let $\mathcal{P}_0$ be the partition of all singletons, i.e., $\mathcal{P}_0 = \{\{1\}, \{2\}, \ldots, \{n\}\}$. Applying T_G to a partition $\mathcal{P}$ containing a non-singleton subset, the probability of moving an element out of a non-singleton subset to a new singleton is

$$\begin{aligned}
&Pr(\text{select a non-singleton element})Pr(\text{move the element to a new singleton}) \\
&= \sum_{\text{i is not a singleton}} g_i \times \frac{1}{\text{number of subsets in } \mathcal{P}} \\
&\geq 2g \times \frac{1}{n} = \frac{2g}{n} > 0,
\end{aligned}$$

where $g = \min\{g_1, \ldots, g_n\} > 0$. So there exists $k(< n)$ such that

$$Pr((T_G)^k(\mathcal{P}) = \mathcal{P}_0) \geq \left(\frac{2g}{n}\right)^k > 0,$$

i.e., $\mathcal{P}$ leads to $\mathcal{P}_0$. From Proposition 4.3.7,

$$\begin{aligned}
Pr((T_G)^k(\mathcal{P}_0) = \mathcal{P}) &= Pr(T_G^k(\mathcal{P}_0) = \mathcal{P}) = Pr(T_G^k(\mathcal{P}) = \mathcal{P}_0) \\
&= Pr((T_G)^k(\mathcal{P}) = \mathcal{P}_0) > 0.
\end{aligned}$$

This implies that any $\mathcal{P} \in \mathcal{A}(\Omega)$ communicates with $\mathcal{P}_0$, which furthers implies that any two partitions in $\mathcal{A}(\Omega)$ communicate. So the Du(n,G,1) process is irreducible.

□

Corollary B.0.3. *The Du(n,G,k) ($k \geq 1$) process is irreducible.*

Proof. Let $\mathcal{P}_0$ be as in Lemma B.0.2, and

$$\begin{aligned}
\mathcal{P}_1 &= \{\{1, 2\}, \{3\}, \ldots, \{n\}\}, \\
\mathcal{P}_2 &= \{\{1, 3\}, \{2\}, \{4\}, \ldots, \{n\}\}.
\end{aligned}$$

It's easy to see that

$$Pr(T_G(\mathcal{P}_0) = \mathcal{P}_1) = Pr(T_G(\mathcal{P}_1) = \mathcal{P}_0) = \frac{g_1 + g_2}{n-1} > 0,$$
$$Pr(T_G(\mathcal{P}_0) = \mathcal{P}_2) = Pr(T_G(\mathcal{P}_2) = \mathcal{P}_0) = \frac{g_1 + g_3}{n-1} > 0,$$
$$Pr(T_G(\mathcal{P}_1) = \mathcal{P}_2) = Pr(T_G(\mathcal{P}_2) = \mathcal{P}_1) = \frac{g_1}{n-1} > 0.$$

This implies that

$$T_G^2 : \mathcal{P}_0 \to \mathcal{P}_1 \to \mathcal{P}_0,$$
$$T_G^3 : \mathcal{P}_0 \to \mathcal{P}_1 \to \mathcal{P}_2 \to \mathcal{P}_0,$$
$$T_G^4 : \mathcal{P}_0 \to \mathcal{P}_1 \to \mathcal{P}_2 \to \mathcal{P}_1 \to \mathcal{P}_0,$$
$$T_G^5 : \mathcal{P}_0 \to \mathcal{P}_1 \to \mathcal{P}_2 \to \mathcal{P}_1 \to \mathcal{P}_2 \to \mathcal{P}_0,$$
$$\ldots.$$

So for any $l \geq 2$,

$$Pr(T_G^l(\mathcal{P}_0) = \mathcal{P}_0) > 0.$$

We have shown in Lemma B.0.2 that for any partition $\mathcal{P} \in \mathcal{A}(\Omega)$, there exists $m \in \mathbb{N}_+$ such that

$$Pr(T_G^m(\mathcal{P}) = \mathcal{P}_0) > 0. \tag{B.1}$$

Let $j \in \mathbb{N}_+ \ni jk - m \geq 2$, we have

$$Pr(T_G^{jk}(\mathcal{P}) = \mathcal{P}_0) \geq Pr(T_G^m(\mathcal{P}) = \mathcal{P}_0) Pr(T_G^{jk-m}(\mathcal{P}_0) = \mathcal{P}_0) > 0. \tag{B.2}$$

From Proposition 4.3.7, we know

$$Pr((T_G^k)^j(\mathcal{P}_0) = \mathcal{P}) = Pr(T_G^{jk}(\mathcal{P}_0) = \mathcal{P}) = Pr(T_G^{jk}(\mathcal{P}) = \mathcal{P}_0)$$
$$= Pr((T_G^k)^j(\mathcal{P}) = \mathcal{P}_0) > 0.$$

This is means that any partition $\mathcal{P}$ in $\mathcal{A}(\Omega)$ communicates with $\mathcal{P}_0$, which further implies that any two partitions in $\mathcal{A}(\Omega)$ communicate. So the Du(n,G,k) process is irreducible. □

Proof of Proposition 4.3.4

Proof. Let $r_1 \equiv F(k) - F(k-1) > 0$, for some $k \in \mathbb{N}_+$. We've already shown that the Du(n,G,k) process is irreducible, i.e., for any two partitions $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{A}(\Omega)$, $\exists t(\mathcal{P}_1, \mathcal{P}_2) \ni r_2 \equiv Pr((T_G^k)^t(\mathcal{P}_1) = \mathcal{P}_2) > 0$. So

$$Pr((T_G^F)^t(\mathcal{P}_1) = \mathcal{P}_2) \geq \left(Pr(T_G^F = T_G^k)\right)^t Pr((T_G^k)^t(\mathcal{P}_1) = \mathcal{P}_2) = r_1^t r_2 > 0,$$

which means that the Du(n,G,F) process is irreducible. □

Proof of Proposition 4.3.5

Proof. Since Ω is a finite set with n elements, $\mathcal{A}(\Omega)$ is finite, the Du(n,G,F) process has a finite state space. We have already shown (Proposition 4.3.4) that the Du(n,G,F) process is irreducible, so the Du(n,G,F) process is positive recurrent (see Paul G. Hoel and Stone, 1972, Corollary 2, page 62), i.e., Proposition 4.3.5 holds. □

Proof of Proposition 4.3.6

Proof. Let $\mathcal{P}_0$ be the same as in Lemma B.0.2, and $r \equiv F(k) - F(k-1) > 0$ for some $k \in \mathbb{N}_+$. As we have shown in Corollary B.0.3,

$$r_2 \equiv Pr((T_G^k)^2(\mathcal{P}_0) = \mathcal{P}_0) = Pr(T_G^{2k}(\mathcal{P}_0) = \mathcal{P}_0) > 0,$$
$$r_3 \equiv Pr((T_G^k)^3(\mathcal{P}_0) = \mathcal{P}_0) = Pr(T_G^{3k}(\mathcal{P}_0) = \mathcal{P}_0) > 0.$$

So we have

$$Pr((T_G^F)^2(\mathcal{P}_0) = \mathcal{P}_0) \geq \left(Pr(T_G^F = T_G^k)\right)^2 Pr((T_G^k)^2(\mathcal{P}_0) = \mathcal{P}_0) = r^2 r_2 > 0,$$
$$Pr((T_G^F)^3(\mathcal{P}_0) = \mathcal{P}_0) \geq \left(Pr(T_G^F = T_G^k)\right)^3 Pr((T_G^k)^3(\mathcal{P}_0) = \mathcal{P}_0) = r^3 r_3 > 0.$$

Thus the period of the Du(n,G,F) process is the greatest common divisor of 2 and 3 which is 1, i.e., the Du(n,G,F) process is aperiodic. □

BIBLIOGRAPHY

Bachmann, F., Hielscher, R., Jupp, P. E., Pantleon, W., Schaeben, H., and Wegert, E. (2010). Inferential statistics of electron backscatter diffraction data from within individual crystalline grains. *Journal of Applied Crystallography*, 43(6):1338–1355.

Basson, F. (1997). Probabilities of random disorientation axes in cubic polycrystals. *Journal of applied crystallography*, 30(2):102–106.

Basson, F. (1998). Response to morawiec's (1998) comment on probabilities of random disorientation axes in cubic polycrystals. *Journal of Applied Crystallography*, 31(3):486–487.

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 2(6):1201–1225.

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2009a). Modeling and inference for measured crystal orientations and a tractable class of symmetric distributions for rotations in three dimensions. *Journal of the American Statistical Association*, 104:1385–1397.

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2010). Finite-sample investigation of likelihood and bayes inference for the symmetric von mises-fisher distribution. *Computational Statistics and Data Analysis*, 54(5):1317–1327.

Bingham, M. A., Vardeman, S. B., and Nordman, D. J. (2009b). Bayes one-sample and one-way random effects analyses for 3-d orientations with applications to materials science. *Bayesian Analysis*, 4:607–630.

Bunge, H. J. (1982). *Texture Analysis in Materials Science*. Butterworths, London.

Camano-Garcia, G. (2006). *Statistics on Stiefel Manifolds*. PhD thesis, Iowa State University, Ames, IA.

Chang, T. and Bingham, C. (1996). Bayesian spherical regression. Bayesian Analysis in Statistics and Econometrics. John Wiley & Sons, New York.

Chang, T. and Rivest, L.-P. (2001). M-estimation for location and regression parameters in group models: A case study using stiefel manifolds. *The Annals of Statistics*, 29(3):784–814.

Chikuse, Y. (2003). *Statistics on Special Manifolds*. Springer, New York.

Downs, T. D. (1972). Orientation statistics. *Biometrika*, 59:665–676.

Du, C., Nordman, D., and Vardeman, S. (2013). Bayesian inference for a new class of distributions on equivalence classes of 3-d orientations defined by crystallographic symmetries. *Technometrics*, Tentatively accepted.

Dubey, A., Hwang, S., Rangel, C., Rasmussen, C., Ghahramani, Z., and Wild, D. (2004). Clustering protein sequence and structure space with infinite gaussian mixture models. In *Pacific Symposium on Biocomputing*, volume 9, pages 399–410. World Scientific Publishing: Singapore.

Gelman, C. A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall, Boca Raton.

Handscomb, D. C. (1958). On the random disorientation of two cubes. *Canadian Journal of Mathematics*, 10:85–88.

Hastie, T. J., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition.

Hielscher, R., Schaeben, H., and Siemes, H. (2010). Orientation distribution within a single hematite crystal. *Mathematical Geosciences*, 42(4):359–375.

Jain, S. and Neal, R. (2004). A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182.

Jupp, P. E. and Mardia, K. V. (1979). Maximum likelihood estimators for the matrix von mises-fisher and bingham distributions. *The Annals of Statistics*, 7(3):599–606.

Khatri, C. G. and Mardia, K. V. (1977). The von mises-fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):95–106.

Kunze, K. and Schaeben, H. (2004). The bingham distribution of quaternions and its spherical radon transform in texture analysis. *Mathematical Geology*, 36:917–943.

Kunze, K. and Schaeben, H. (2005). Ideal patterns of crystallographic preferred orientation and their representation by the von mises - fisher matrix or bingham quaternion distribution. *Materials Science Forum*, 495-497:295–300.

Lau, J. and Green, P. (2007). Bayesian model-based clustering procedures. *Journal of Computational and Graphical Statistics*, 16(3):526–558.

León, C. A., Massé, J.-C., and Rivest, L.-P. (2006). A statistical model for random rotations. *Journal of Multivariate Analysis*, 97:412–430.

Mackenzie, J. K. (1958). Second paper on statistics associated with the random disorientation of cubes. *Biometrika*, 45(1-2):229–240.

Mackenzie, J. K. and Thomson, M. J. (1957). Some statistics associated with the random disorientation of cubes. *Biometrika*, 44(1-2):205–210.

Mardia, K. and Jupp, P. (2000). *Directional Statistics*, volume 494. Wiley, New York.

Matthies, S., Muller, J., and Vinel, G. W. (1988). On the normal distribution in the orientation space. *Textures and Microstructures*, 10:77–96.

Matthies, S., Vinel, G. W., and Helming, K. (1987). *Standard Distribution in Texture Analysis: Maps for the Case of Cubic - Orthorhombic Symmetry*, volume 1. Akademie, Berlin.

Morawiec, A. (1995). Misorientation-angle distribution of randomly oriented symmetric objects. *Journal of applied crystallography*, 28(3):289–293.

Morawiec, A. (1996). Distributions of rotation axes for randomly oriented symmetric objects. *Journal of applied crystallography*, 29(2):164–169.

Morawiec, A. (1998). Comment on probabilities of random disorientation axes in cubic polycrystals by basson. *Journal of Applied Crystallography*, 31(3):486–486.

Morawiec, A. (2004). *Orientations and rotations: computations in crystallographic textures*. Springer, Berlin.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Nobile, A. and Fearnside, A. (2007). Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Statistics and Computing*, 17(2):147–162.

Paul G. Hoel, S. C. P. and Stone, C. J. (1972). *Introduction to Stochastic Processes*. Boston, Houghton Mifflin.

Pio, R. (1966). Euler angle transformations. *Automatic Control, IEEE Transactions on*, 11(4):707–715.

Pitman, J. and Yor, M. (1997). The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25(2):855–900.

Prentice, M. J. (1986). Orientation statistics without parametric assumptions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):214–222.

Qiu, Y., Nordman, D., and Vardeman, S. (2014). One-sample bayes inference for existing symmetric distributions on 3-d rotations. *Computational Statistics and Data Analysis*. to appear.

Rancourt, D., Rivest, L.-P., and Asselin, J. (2000). Using orientation statistics to investigate variations in human kinematics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49(1):81–94.

Randle, V. (2003). *Microtexture Determination and Its Aplications*. Maney for the Institute of Materials, Minerals and Mining, London.

Savyolova, T. I. and Nikolayev, D. I. (1995). Normal distribution on the rotation group $so(3)$. *Textures and Microstructures*, 29:201–233.

Watson, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

Zhou, W. and Wang, Z. L. (2007). *Scanning Microscopy for Nanotechnology: Techniques and Applications*. Springer, New York.